

生命科学实验指南系列



Protein Engineering Protocols

现代蛋白质工程实验指南

［德］ K.M. 阿恩特 K.M. 米勒 编著
苏晓东 曾宗浩 杨 娜 译
苏晓东 高 嵘 校

科学出版社

生命科学实验指南系列 · 典藏版



图解微生物实验指南
免疫学技术及其应用
生物衰老：研究方法与实验方案
精编细胞生物学实验指南
植物蛋白质组学实验指南
蛋白质纯化指南（原书第二版）
环境基因组学实验指南
实验动物血液生理生化参考手册
生理学实验指南
精编免疫学实验指南
酵母遗传学方法实验指南
人干细胞培养
抗体制备及使用实验指南
病毒的电子显微学研究
植物生物学与生态学实验
神经生物学实验原理与技术
DNA微阵列实验指南
基因转移：DNA和RNA的转运与表达
生物实验室管理手册（原书第二版）
精编人类遗传学实验指南
单分子技术实验指南
现代蛋白质工程实验指南
活细胞成像（原书第二版）
遗传变异分析实验指南
表皮细胞实验指南
分子克隆实验指南（原书第三版）（上下册）
精编分子生物学实验指南（原书第五版）
现代神经科学研究技术
生命科学实验设计指南
现代生物化学与分子生物学仪器与设备
分子细胞遗传学——技术和应用
精编蛋白质科学实验指南
实验细胞资源的描述标准与管理规范
实验动物设施运行管理指南
元基因组学：方法和步骤（影印版）
现代工业微生物学实验技术
真核生物转录调控——概念策略与技术（原书第二版）
动物细胞培养——基本技术和特殊应用指南（原书第六版）



科学出版中心 生物分社
联系电话：010-64012501
E-mail：lifescience@mail.sciencep.com
网址：http://www.lifescience.com.cn

销售分类建议：分子生物技术



赛拉艾芙
生命科学订阅号


ISBN 978-7-03-047486-5
9 787030 474865

定价（全套）：4500.00元

生命科学实验指南系列·典藏版

现代蛋白质工程实验指南

Protein Engineering Protocols

〔德〕 K. M. 阿恩特
K. M. 米勒 编著
苏晓东 曾宗浩 杨 娜 译
苏晓东 高 嵘 校

科学出版社
北 京

图字：01-2008-3816 号

内 容 简 介

“生命科学实验指南系列”图书均出自名家，包括众多从 Cold Spring Harbor Laboratory Press 和 John Wiley & Sons 等国际知名出版社引进的实验室必备工具书，是生命科学领域最先进、实用、权威的实验手册类优秀图书。该系列图书简单明了，囊括了全世界最著名的生物类实验室操作方法，无论是初学者还是需要深入研究的科研工作者都能从中获益。该系列图书在读者群中有较高的知名度和美誉度，特别是以《分子克隆实验指南》和《精编分子生物学实验指南》为代表，堪称经典，分别被喻为生命科学领域的“蓝宝书”和“红宝书”。现挑选其中的精品集结成典藏版。

图书在版编目（CIP）数据

生命科学实验指南系列：典藏版/雷东锋等编著. —北京：科学出版社，2016
ISBN 978-7-03-047486-5
Ⅰ.①生… Ⅱ.①雷… Ⅲ. ①生命科学–实验–指南 Ⅳ.①Q1-0
中国版本图书馆 CIP 数据核字(2016)第 043878 号

责任编辑：王 静 李 悦
责任印制：张 伟 / 封面设计：刘新新

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
北京厚诚则铭印刷科技有限公司 印刷
科学出版社发行 各地新华书店经销
*
2016 年 7 月第 一 版 开本：787×1092 1/16
2016 年 7 月第一次印刷 印张：1310 1/2
字数：31 074 000
定价：4500.00 元
(如有印装质量问题，我社负责调换)

译 者 序

自 20 世纪六七十年代分子生物学及重组蛋白质技术建立及普遍应用以来，生物技术已经取得了非常广泛且令人瞩目的成果，“蛋白质工程”也似乎渐渐变成了一个古老的话题。然而，近二十年来，蛋白质工程的相关技术取得了长足的进展，特别是在蛋白质的定向进化（directed evolution）及理性设计（rational design）等方面。一年多以前，我们有幸得到英国剑桥大学 John M. Walker 教授（他因阐明了 ATP 合成酶的结构与机理而荣获 1997 年诺贝尔化学奖）主编的《分子生物学方法》*Methods in Molecular Biology* 系列丛书之一，由德国弗莱堡大学的 Katja M. Arndt 和 Kristian M. Müller 编著的 *Protein Engineering Protocols* 一书，此书由欧美及日本的多位蛋白质化学及分子生物学相关领域国际知名专家撰写，从理论到实践较为详细地介绍了现代蛋白质工程这两个最新前沿及其相关进展实例。

所谓定向进化，是指对蛋白质（通常是通过其编码核酸的随机合成及重组）引入大量的随机突变体，然后快速高效地选择或筛选具有相应特性的突变体蛋白的过程，定向进化可以通过体外或者体内实验进行，本书主要介绍的是体外定向进化方法。人们通过多轮的突变和筛选取得特定性质的突变体蛋白，可以说定向进化方法是在模拟自然界的进化及自然选择过程，只不过我们的选择压力是人为在实验室中施加的。另外，我们还可以模拟自然界有性生殖中的重组过程，通过各种 DNA 混编（DNA shuffling）技术将筛选得到的突变体进一步混合匹配以得到更优化的结果。一般说来，定向进化不需要预先知道蛋白质的详细三维结构信息，尽管这些信息有时对于设计定向进化实验很重要，也不需要预测一个突变体对蛋白质会产生怎样的功能及稳定性影响。事实上，很多定向进化的实验结果表明目标蛋白质特性的产生往往是由意想不到的突变造成的，通过定向进化实验我们能够学习到很多关于蛋白质结构与功能的新知识。

尽管目前蛋白质的折叠问题仍然没有取得根本性突破，我们还不能够一般地预测任意序列的蛋白质折叠类型及其三维结构，然而，随着结构基因组学及相关计算生物学的发展，蛋白质三维结构研究与预测已经取得了重大进展，截至 2010 年底，PDB 数据库（http://www.rcsb.org/pdb/home/home.do）已经收录了七万多个蛋白质三维结构坐标，并且统计数据表明从 2008 年以来，PDB 库中就没有收录到新的蛋白质折叠类型，这意味着当前 PDB 数据库已经覆盖了绝大部分已知蛋白质的折叠类型。在此基础上，根据目前已知蛋白质的三维结构与功能细节合理化设计新的蛋白质已经成为可能，本书介绍了这一方面的理论基础及应用实例。

本书在理论上从蛋白质工程中的设计计算策略和进化策略出发，讨论了采用计算方法的组合式蛋白质设计策略及蛋白质进化库的设计与筛选，并且介绍了一个行之有效的基于极性与非极性氨基酸的二元组图进行蛋白质设计的方法细节。在实践上本书详细介绍了蛋白质中非天然氨基酸的整体掺入方法；作为理性设计实例列举了基于钙调素与荧光蛋白融合的钙指示剂；人工锌指蛋白的设计与合成，以及基于纤连蛋白类型Ⅲ结构

域框架的抗体模拟等。在定向进化方法的应用方面，介绍了噬菌体展示技术、核糖体失活展示系统、利用核苷酸交换和剪切技术进行 DNA 碎裂和定向进化、简并寡核苷酸基因混编，以及其他酶的定向进化新方法。我们很高兴受科学出版社的委托将此书翻译成中文，介绍给我国广大蛋白质科研工作者，希望本书的中译本对于推动我国现代蛋白质工程学科的进一步发展做出贡献。

我们衷心感谢参与本书翻译及校正工作的很多同仁，他们包括中国科学院生物物理研究所的王大成院士，他和科学出版社编辑一起向我们推荐并介绍了此书。还要感谢北京大学生命科学学院的李兰芬老师，王娟、刘晓艳、雷剑博士，以及傅天民、王子曦及金坚石等同学。

译者

2011 年 2 月

前 言

蛋白质工程是将分子生物学、蛋白质结构分析、理论计算以及生物化学有机结合的学科,其目标是发展实用的或有价值的蛋白质。本书涉及蛋白质工程领域中两个普遍的但不互相排斥的策略。第一个策略是理性设计(rational design),科学家根据蛋白质的结构与功能细节设计相应的蛋白质;第二个策略是所谓定向进化(directed evolution),通过对蛋白质引入随机突变,然后选择或筛选具有相应特性的突变体。通过多轮突变和筛选的方法,可以说是模拟自然界的进化过程。另外,通过 DNA 混编(DNA shuffling)技术将筛选得到的突变体进一步混合匹配以得到更优的结果,从而模拟自然界有性生殖中的重组过程。

本书的第一部分介绍了蛋白质理性设计策略,包括理论计算方法,通过引入非天然氨基酸来扩展生物学字母表,以及一些很有说服力的例证说明所设计蛋白质的全新特性。尽管引入突变的策略已经成为常规,但预测和推断这些突变产生的效果仍是非常具有挑战性的,除了基本的蛋白质结构信息外,还需要对整个系统有深刻的理解。因此,这一部分主要讲述如何设计具有目标特性的蛋白质,并挑选了涵盖蛋白质工程大部分技术的实例,如蛋白质-蛋白质相互作用、DNA 结合、抗体模拟,以及酶活性设计等。

本书的第二部分主要讲述了进化技术。与理性设计不同的是,定向进化策略不需要预先知道蛋白质的三维结构信息,也不需要预测一个突变对蛋白质会产生怎样的影响。事实上,定向进化的实验结果表明目标特性的产生往往是由意想不到的突变造成的。定向进化策略成败的关键因素有:进化库的设计与质量、进化及 DNA 重组方法的选择,以及筛选方法的选择。因此,第二部分主要介绍了上述几个方面的内容,包括进化库设计的一般方法、进化库质量的统计评估、DNA 混编的新方法,以及不同的选择筛选策略等。同时也给出了不同特性进化的一些实例,如蛋白质折叠类型、热稳定性,以及酶活性等。

本书全面地介绍了蛋白质工程各个阶段所使用的方法,综合了完备的理论基础和具体的实验细节,适合该领域各个层次的科研工作者使用。感谢参与此书编著的所有人员作出的卓越贡献,特别是从书系列主编 John M. Walker 教授(1996 年诺贝尔化学奖得主)在本书编撰过程中给予的耐心指导和帮助。

K. M. 阿恩特

K. M. 米勒

原作者及其单位

Katja M. Arndt • *Institut* für *Biologie III, Universität Freiburg, Freiburg, Germany*

Jamie M. Bacher • *The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA*

Peter L. Bergquist • *Biotechnology Research Institute, Macquarie University, Sydney, NSW, Australia; Department of Molecular Medicine and Pathology, Auckland University Medical School, Auckland, New Zealand*

Luke H. Bradley • *Department of Chemistry, Princeton University, Princeton, NJ*

François-Xavier Campbell-Valois • Dé*partement de Biochimie,* Université *de* Montréal, Montréal, Québec, *Canada*

Michel Denault • *Department of Quantitative Methods, HEC* Montréal, Montréal, Québec, *Canada*

Andrew D. Ellington • *Institute for Cellular and Molecular Biology and Department of Biochemistry, University of Texas, Austin, TX*

Birte K. Feld • *Department of Chemistry and the Institute for Genomics and Bioinformatics, University of California, Irvine, CA*

Peter Friedhoff • *Institut* für *Biochemie, Justus-Liebig-Universität, Giessen, Germany*

Satoshi Fujita • *Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Hongo, Tokyo, Japan; Research Institute for Cell Engineering, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan*

Farid J. Ghadessy • *MRC Laboratory of Molecular Biology, Cambridge, United Kingdom; Department of Oncology, University College Medical School, London, United Kingdom*

Moreland D. Gibbs • *Biotechnology Research Institute, Macquarie University, Sydney, NSW, Australia*

Michael H. Hecht • *Department of Chemistry, Princeton University, Princeton, NJ*

Jochen Hecky • *Institut* für *Biology III, Universität Freiburg, Freiburg, Germany*

Philipp Holliger • *MRC Laboratory of Molecular Biology, Cambridge, United Kingdom*

Mitsuhiko Ikura • *Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada*

Akiko Koide • *Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL*

SHOHEI KOIDE · *Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL*

HIDETOSHI KONO · *Computational Biology Group, Neutron Science Research Center, Quantum Beam Science Directorate, Japan Atomic Energy Agency, Kyoto, Japan*

JODY M. MASON · *Institut* für *Biologie III, Universität Freiburg, Freiburg, Germany*

STEPHEN W. MICHNICK · Dé*partement de Biochimie*, Université *de* Montréal, Montréal, Québec, *Canada*

KRISTIAN M. MÜLLER · *Institut* für *Biologie III, Universität Freiburg, Freiburg, Germany*

ATSUSHI MIYAWAKI · *Laboratory for Cell Function and Dynamics, Advanced Technology Development Center, Brain Science Institute, RIKEN, Wako City, Saitama, Japan*

WATARU NOMURA · *Institute for Chemical Research, Kyoto University, Uji, Kyoto, Japan*

JOELLE N. PELLETIER · Dé*partement de Chimie*, Université *de* Montréal, Montréal, Québec, *Canada*

ALFRED PINGOUD · *Institut* für *Biochemie, Justus-Liebig-Universität, Giessen, Germany*

JEFFERY G. SAVEN · *Makineni Theoretical Laboratories, Department of Chemistry, University of Pennsylvania, Philadelphia, PA*

ASAKO SAWANO · *Laboratory for Cell Function and Dynamics, Advanced Technology Development Center, Brain Science Institute, RIKEN, Wako City, Saitama, Japan; Brain Science Research Division, Brain Science and Life Technology Research Foundation, Itabashi, Tokyo, Japan*

SACHDEV S. SIDHU · *Department of Protein Engineering, Genentech Inc., South San Francisco, CA*

SABINE C. STEBEL · *Institut* für *Biologie III, Universität Freiburg, Freiburg, Germany*

YUKIO SUGIURA · *Faculty of Pharmaceutical Sciences, Doshisha Women's University, Koudo, Kyotanabe, Japan*

KAZUNARI TAIRA · *Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Hongo, Tokyo, Japan; Gene Function Research Laboratory, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Science City, Japan*

PETER THUMFORT · *Department of Chemistry, Princeton University, Princeton, NJ*

KEVIN TRUONG · *Institute of Biomaterials and Biomedical Engineering, Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada*

WEI WANG · *Makineni Theoretical Laboratories, Department of Chemistry, University of Pennsylvania, Philadelphia, PA*

YINAN WEI · *Department of Chemistry, Princeton University, Princeton, NJ*

GREGORY A. WEISS · *Department of Chemistry and the Institute for Genomics and Bioinformatics, University of California, Irvine, CA*

CHRISTINE WURTH · *Department of Chemistry, Princeton University, Princeton, NJ*
JING-MIN ZHOU · *Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Hongo, Tokyo, Japan; Gene Function Research Laboratory, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Science City, Japan*

目　录

第一部分　蛋白质工程中的设计与计算策略

1 采用计算方法的组合式蛋白质设计策略

Hidetoshi Kono, Wei Wang, and Jeffery G. Saven

概要 计算方法一直在蛋白质设计中发挥重要作用。本工作主要集中在搜索蛋白质序列空间，以找到一条或数条与已知结构和功能相容的蛋白质序列。在期望的功能和结构限制下，概率性计算方法为所容许的氨基酸变化范围提供信息。这样的方法可用于指导建立蛋白质的单个序列或组合库。

关键词 全新蛋白质设计；组合库；蛋白质计算设计；偏好编码

1.1 介 绍

1.1.1 蛋白质设计

通过对蛋白质结构（包括那些具有特殊功能结构）的设计，研究者可以增进对决定蛋白质折叠状态特征的力和效应的理解。另外，对特定折叠结构设计的控制，可能得到新的具有生物效能和特异性的合成蛋白。这样的应用包括在医药、传感器、催化剂和材料等领域。甚至在尚未完全和定量地理解决定蛋白质结构的力的条件下，蛋白质设计仍有可能取得成功。

但是，由于决定蛋白质折叠态的相互作用的复杂性和精细性，蛋白质设计不是一件轻而易举的事。蛋白质是巨大的分子（含数十到数百个氨基酸残基），折叠态的给定需要许多的结构变量，其中包括序列、主链拓扑和侧链构象。即使主链结构已经给定，每个残基仍可能有多重构象。除了结构的复杂性之外，还有序列的复杂性。设计意味着从无数的可能的序列中，鉴别出可折叠的序列。在折叠蛋白中观察到的高度“一致性”引导这个搜索过程[1]。一般来说，处于折叠状态的蛋白质，在原子水平上，通过有利的范德华相互作用、疏水残基与溶剂隔离，并使大多数氢键作用都被满足而恰当地堆积。但这种一致性通常都是复杂的，可能没有什么能使问题简化的对称性。另外，精确地定量非共价相互作用属于最困难的一类问题，并且，估计残基替换或结构有序化的自由能，仍然是计算研究中最深奥的领域[2,3]。与预测能力相反，目前，我们还不能期望用详细的模拟估计自由能变化的方法，来确定大量序列的相对稳定性变化。尽管如此，从小分子和蛋白质数据库导出的分子位能，确实包含了已知的，对决定蛋白质结构起重要作用的，相互作用和力的部分信息。在某些情况下，这些位能的优化，已经在蛋白质设计中获得显著的成功[4]。这样的位能肯定是近似的，并且这样设计的任何序列，很可能对特定的位能和采用的目标结构敏感。作为另一种选择，在这些位能中包含的部分信息也可以做概率性的应用，以得到出现某种氨基酸的可能性。概率性方法也适合于确定可折叠到同一结构序列的完整可变性，因为似乎存在大量这样的序列——远

大于可以用序列搜索或列举方式所能处理的。

这样的概率性方法也特别适合于在蛋白质组合实验中的全新设计，这些实验能产生并快速测试许多序列。虽然组合方法能处理大量序列（$10^4 \sim 10^{12}$），但这些数量与可能的蛋白质序列数比较仍然是无穷小，如对 100 个残基的蛋白质，这个数是 $20^{100} \approx 10^{130}$（为了对 10^{130} 这个数有多大有个一般的了解，假定合成出一条 100 个残基的蛋白质序列约需要10 000/ N g 物质，$N \approx 6 \times 10^{23}$，是阿伏伽德罗常数，那么合成出 10^{130} 条序列需要超过 10^{107} kg的物质。这个数量远超过目前我们所知宇宙物质的总质量 10^{35} kg——译者注）。于是，即使是采用组合法，我们仍然必须集中于序列空间选定的一小部分。通过预先观察，在蛋白质中选定若干残基位点并在这些选定的位点允许残基的全部或部分可变性（全部可变性，即可以替换为 20 种氨基酸中任何残基；部分可变性，只替换为某些类型的残基——译者注）来实现对序列空间的限定。近来发展了可以在宽得多的范围内追踪序列可变性，并提供扫视和聚焦序列空间的定量计算方法。在这里，我们讨论序列设计的计算方法，把重点放在处理给定结构位点特异氨基酸可变性的概率方法。

1.1.2　蛋白质设计的定向方法

这里的“定向蛋白质设计”是指鉴定出一条（或一组）可能折叠为预先指定的主链结构的序列。然后可以用多肽合成或基因表达的办法，实验性地实现每一条这样的序列，以确认其折叠态及其他分子性质。早期的设计努力，是在已观察到的自然发生的结构和已被确认的蛋白质序列的指导下完成的，它们有着重要的二级结构，但并不必有明确的三级结构[5]。由于能够定量化并以表格列出残基间相互作用，计算方法已经极大地加速了蛋白质设计的成功率。典型情况下，这样的方法使序列搜索成为优化过程，在过程中改变氨基酸身份和侧链构象，以优化定量化序列结构相容性的打分函数，对所有 m^N 个可能序列的完全搜索，只有仅仅少量的残基 N 是允许改变的，或允许改变的氨基酸数目显著地减少，如从 $m=20$ 减少到 $m=2$ 时才是可行的。为了达到从内部平均来看有利的原子间相互作用的合理堆积序列，必须搜索每个氨基酸的不同侧链构象（旋转异构态）（见参考文献 [6]）。结果，因为一个残基所可能有的状态数 m 增加 10 倍或更多——这取决于每个残基的旋转异构态数目和增加搜索的复杂性。如果只有很少的残基被允许改变，并且剩下的残基的构象受到限制，就可以对所有可能的组合完成完整数值计算，以鉴别出低能量的序列旋转异构态组合。由于（这个组合数）对链长和旋转异构态数目的指数依赖性，这样的完整数值计算在典型的情况下不能实现。在这样的情况下，序列空间可以用定向的方式取样以逐渐朝优化的（或部分优化的）序列方向移动。随机方法，如遗传算法和模拟退火，包含对序列空间的部分随机式搜索。在这种搜索中，搜索逐渐地移向高分（低能）序列[7~10]。这样的搜索有足够的“噪声”或重组，以允许越过序列-旋转异构地形图的局部极小值。当运用于精细到原子的表象中时，随机方法基本上集中于用疏水残基重新堆积结构的内部[9]，并已被用于 434 Cro[10]、泛素[11]、G 蛋白的 B1 结构域[12]、WW 结构域[4]和螺旋束[13,14]的野生型结构。虽然在许多情况下这些方法对鉴别出实验可行的序列[4,15]有帮助，但是随机搜索法不必鉴别出整体优化点[16]。对只含有位点和对相互作用的位能，通过排除法，如“死点排除法”，能找到整体优化点[16~20]。这样的方法可连续地移除不可能是整体优化点的氨基

酸-旋转异构态，直到再没有态可被移除。Mayo 研究组应用这一方法，已使拟 28 残基锌指蛋白[21]和在疏水和极性位点模式化之后的 51 残基同源域蛋白模体[22]的完整序列设计自动化。该组还重新设计了数种蛋白质内的部分残基组[23~25]。其功能性质，如结合金属或催化，也可以包括作为设计过程的元素[26~28]。蛋白质定向设计的要素和算法是最近一些综述的主体[4,29,30]。

尽管取得了某些惊人的成功，但序列定向设计的计算方法在鉴别折叠为特定结构的蛋白质序列特征上仍有局限性。随机方法可用于大蛋白，也容许多位点上的同时变化，但是，即使用于小蛋白，这样的计算消耗的机时和资源也是巨大的。定向方法对于使用的能量或打分函数必定是敏感的，因为它能鉴定能量函数的优化点。但是，所有这样的函数也必定是近似的，并且能量函数中的不确定因素可能不允许对整体优化点的搜索。许多天然存在的蛋白质是没有优化的。实际上，多数蛋白质只有微小的折叠稳定性，如 $\Delta G^{\circ}<10$kcal①/mol[31]。更有甚者，在功能上与其他分子结合的序列在结构稳定性上不必是整体优化的。重要的是发展与定向蛋白设计互补的方法，这些方法揭示可能折叠为某特定结构但又可能在结构上有未被优化的序列的特征。这样的技术可用于设计蛋白质序列。另外，这样的计算方法还可用于新型的蛋白质设计研究——组合式实验，即实验中大量蛋白质可以同时合成和筛选。

1.1.3 蛋白质设计的概率性方法

在蛋白质涉及的范围内，我们用定点氨基酸概率而不是特定的序列来描述“概率性蛋白质设计”。相对于定向的或决定论方法，概率性方法是常用于对问题只有部分信息场合的定量科学。对蛋白质设计，折叠过程的复杂性和不确定性促成了这样的概率性方法。蛋白质折叠是一个复杂的动态过程，有无数的相互作用规定折叠状态。每一个导致稳定的非共价键相互作用在大小上都是可以相互比较的，似乎没有哪一个具有压倒优势，以致于在折叠中起决定作用。定量化这些相互作用的办法，必定是近似的（见注1）。概率性设计方法也直接提供非常有用的序列信息，特别是在结构上重要的氨基酸。氨基酸概率可以引导特定序列的设计，也能够凸显能容忍突变、对结构只有微小影响的位点；在几轮蛋白质设计之后，这样的位点可以成为用来改变的目标。

概率性方法可以以几种方式应用于蛋白质设计。序列应该以符合计算出的概率方式生成。首先，最直接的选择是一个共用序列，或在每一个位点用最可能的氨基酸组成的序列。在必要时，可以重复地计算，逐次地增加蛋白质中（已经确定的）的残基。用这样的方法，已得到 114 个残基的双核金属蛋白[32]和一个完整膜蛋白的可溶性变体[33]。其次，计算概率可用于引导对序列的搜索，已提出基于 Monte Carlo 的方法。在 Monte Carlo 轨道的每一点决定序列的接受或拒绝时，计算的氨基酸概率用作有倾向性的选择标准[34]。用这样的方法处理相关的氨基酸身份，但要付出用于搜索的计算运转开销，如果有信息可用于搜索，开销可以减少。最后，概率性方法可以用来定量地指导蛋白质组合库的设计[35]。

1.1.4 组合实验

组合的蛋白质实验可以用来研究序列结构相容性和发现折叠为特定结构的新序列。

① 1cal＝4.1868J，后同。

在蛋白质组合设计实验中，筛选大量的序列（实物库），以找到以折叠为预先确定结构的迹象。取决于序列的离散性是如何产生和检验的，这类实验可以探测大量序列序列数可高达 10^{12}[36]。可以用选择性分析，如配体结合或催化活性筛选序列（实物）库。以序列离散程度受研究者控制的方式，这样的实验可以“超越蛋白质序列数据库”。由于去掉了与天然蛋白进化压力的耦合，可以研究对折叠（和其他生物学性质）重要的特征。组合方法已用于鉴别螺旋蛋白[37~39]、泛素变体[40]、单层自组装蛋白[41]、具有纤维样性质的蛋白[42]以及稳定的寡聚螺旋[43]。最近发表了几篇出色的关于组合实验和方法的综述[44~47]。

1.2　方　　法

蛋白质设计中的概率性方法是提供在一个特定的蛋白质结构中对某氨基酸（出现）在一指定位点的概率估计。这里，我们讨论几种估计这些概率的方法并将重点放在直接解出这些概率的基于熵的自洽公式。

1.2.1　关联序列的比对

蛋白质结构的序列可变性可用序列和结构数据库来探讨。已知折叠为非常相似结构的序列可以从蛋白质数据库或结构比对数据库中鉴别出来[48]。如果一个序列的结构已知，具有足够序列相似性的［如序列同一性（identity）大于 40%］可以认为共享同一结构。对这样结构相似蛋白质的多序列比对可以把氨基酸位点特异的概率简单地估计为比对中每一位置（出现）每一氨基酸的频度[49]。这样一组概率常称为序列剖面（sequence profile）。如果序列的数量不够，以致于某些氨基酸在某些特别位点上从未出现过，伪计数（psedocount）和其他方法可以用来规整化这些频度，以使它们在折叠为选定结构这点上更有代表性[50]。虽然如此，从这样的剖面得到的概率将使数据库中序列的性质产生严重偏差。因为存在大量相似性很低的序列折叠类似结构的例子，我们希望在更广的范围内对序列可变性得到完整的理解。从数据库导出的剖面也不适合于设计数据库中没有序列的新蛋白质结构。用一个给定的主链结构为模板，更普遍的计算方法可从头确定氨基酸概率。

1.2.2　建立剖面的定向搜索方法

定向搜索方法的重复应用可以估计一个序列整体的性质。对这类计算，通过给定主链原子坐标来选定一个目标结构。如果采用单个的蛋白质结构，几个最新的直接设计研究得到了与野生型序列相当相似的序列[51~54]。对一个给定结构，可以独立运行多序列搜索计算，以得到一组序列，这些序列的比对产生位点特异概率。Desjarlais 及其合作者，对与一特定折叠一致性的极相关的结构系统的每一个成员独立地运行了他们的序列预测算法[55]。对每一个结构，鉴别出一个优化的“成核”序列，随后，对整个结构探索序列/旋转异构。这个方法已被用于鉴定与小 β 片 WW 结构域[4,55]折叠相容的序列。对一特定折叠的 100 个微结构变体（1Å 均方差）的每一个，应用序列预测算法构建了比应用单个结构更为离散得多的计算剖面[56]。Xencor Inc. 的工作者对一优化序列（其中 β 内酰胺酶活性位点附近的残基被替换）进行了采样，采用了 Monte Carlo 采样

法[57]，找到了对一种抗生素的抗性增加了 1000 倍多的序列。但是，构建剖面的这些方法非常耗费计算资源。因为，为建立氨基酸位点特异的频度，要完成重复的定向搜索。

1.2.3 序列系综的统计理论

已经建立起统计的、基于熵的公式，对给定的主链结构鉴定出一组位点特异的氨基酸概率，而不只是最优的序列[58,59]。来源于统计学的理论被用来处理与主链结构相容的序列的数目和构成。这一理论也处理构成适合的整个空间，而不只是对实验和数值计算及取样可达到的小部分空间。亚优化序列的特性很容易检验。大蛋白结构（多于 100 个残基）计算起来很容易。这里的“熵”是指与目标结构相容的序列数。这个来源于热力学的概念被用来减少可能的序列数：对序列的限制减少了熵，并且伴随能量的降低熵也在减少。

方法中的输入是目标主链结构，以及定量化序列-结构相容性的能量函数。对于一个目标主链结构，该方法产生每一个氨基酸（出现）在每一个残基位点的概率（见注 2）。在理论中整体的特性（如序列在该目标中的总能量）和局部特性（如在某特定位点所容许的氨基酸），两者可以作为限制包含在方法中。许多氨基酸概率的集合是可能的。这个方法用极大化有效熵的方法确定“最可几”（“最可能”）的这样的集合，借此，这种极大化是受限的。此方法有效地通过这样的限制来为系统提供手段，以减小需要搜索的序列空间体积，达到实验可及的水平。

在限制函数规定的具有期望的特性序列中，令 w_i［α，r_k（α）］表示氨基酸 α 出现在残基位置 i 并使得其侧链为一组离散构象——r_k（α）（旋转异构态；参考［6］和［60］）中的任何一个的概率。总的序列-构象熵——S_c（此处简单地称为“构象熵”）可以定义为

$$S_c = -\sum_{i,\alpha,k} w_i[\alpha, r_k(\alpha)] \ln w_i[\alpha, r_k(\alpha)]$$

求和遍及每一个序列位点 i 和所有可能的氨基酸 α。对每一种氨基酸求和也遍及 k 种可能的旋转异构态——r_k（α）。在限制条件 f_i 之下，通过极大化 S_c 来得到 w_i［α，r_k（α）］。极大化采用拉格朗日乘子法[61]。w_i［α，r_k（α）］的变分泛函 V 定义为

$$V = S - \beta_1 f_1 - \beta_2 f_2 - \cdots$$

一般来说，限制条件 f_i 也是概率 w_i［α，r_k（α）］的函数。在确定与特定的限制相容的状态概率时，第 m 个限制函数 f_m 被限定取值 f_m^o。确定概率的方程组和拉格朗日乘子的形式为（见注 3）：

$$0 = \partial V / \partial w_i[\alpha, r_k(\alpha)]$$

$$f_m^o = f_m\{w_i[\alpha, r_k(\alpha)]\}$$

这个大的耦合非线性方程组用求根法（root-finding）求解。虽然这样的方法有很多选择，我们找到一种可以广泛采用的整体收敛的方法[62]。

1.2.3.1 能量函数

在计算中考虑两种能量——构象能 E_c 和环境能 E_{env}，并在极大化构象熵中用作限制条件。

构象能 E_c 用基于原子的位能——AMBER 力场[63]计算。E_c 包括范德华相互作用，带有与距离相关的介电常数（$4\varepsilon r_{ij}$）的静电相互作用，以及修正后的氢键项[64]。对一

个特定序列（α_1，…，α_N），其中氨基酸的构象态是［r_1（α_1），…，r_N（α_N）］，E_c是

$$E_c=\sum_i \varepsilon_i[\alpha, r_k(\alpha)]+\sum_{i,j>i}\varepsilon_{i,j}[\alpha, r_k(\alpha); \alpha', r_{k'}(\alpha')]$$

在考虑蛋白质能量函数的时候，单体项 $\varepsilon_i[\alpha, r_k(\alpha)]$ 包括主链和侧链原子的相互作用，以及氨基酸的参考能量（见 1.2.3.3 小节）。双体项 $\varepsilon_{i,j}[\alpha, r_k(\alpha); \alpha', r_{k'}(\alpha')]$ 为对结构中两个不同位点的两个旋转异构态间相互作用的求和。对享有共同能量特性的大量序列，我们假定 E_c 由于序列改变引起的围绕其平均值的涨落不大。那么，我们可以写出

$$E_c\approx \bar{E}_c=\sum_{i,\alpha,k}\varepsilon_i[\alpha, r_k(\alpha)]\,w_i[\alpha, r_k(\alpha)]+\sum_{\substack{i,j>i\\ \alpha,\alpha'\\ k,k'}}\varepsilon_{i,j}[\alpha, r_k(\alpha); \alpha', r_{k'}(\alpha')]\,w_i[\alpha, r_k(\alpha)]\,w_j[\alpha', r_{k'}(\alpha')]$$

作为另一个限制项，引入环境能 E_{env} 以在统计理论内用等效的方式计入疏水效应[59]。这个位能考虑了氨基酸的表面暴露倾向。我们可以用氨基酸概率把 E_{env} 写成

$$E_{env}\approx \bar{E}_{env}=\sum_{i,\alpha,k}\varepsilon_{env}[\alpha, r_k(\alpha)]\,w_i[\alpha, r_k(\alpha)]$$

式中，ε_{env} 为在 1.2.3.2 小节定义的局部环境能量。需要注意的是这个能量不包含双体相互作用并且只取决于在每一个位置的氨基酸和旋转异构态。

1.2.3.2　溶解和疏水能

定量化疏水作用和其他溶液特性在蛋白质设计的方法上是一个重要的参数。用计算来检验序列中大量的变化是不切实际的，即使是计算溶液可及表面积，它常常与疏水倾向相关得很好，也可能要消耗大量计算资源。在用于统计计算一致的实用方法考虑溶液效应的努力中，作为每一个位点附近的β碳原子密度 ρ 的函数，引进了环境能[59]。一般来说，疏水残基倾向于定位在蛋白质的掩埋区，而亲水残基则倾向于定位在表面。因而，疏水残基倾向于有比亲水残基更高的β碳原子密度。通过 500 个不同的、已知结构的球蛋白，我们推导了计算氨基酸有效势能的通用“统计”势能方程

$$\varepsilon_{env}(\alpha, \rho)=-T_e\ln\frac{p(\alpha, \rho)}{p(\alpha)\,p(\rho)}$$

式中，$p(\alpha, \rho)$ 为观察到残基 α 的局部β碳原子密度为 ρ 的次数；$p(\alpha)$ 为观察到残基 α 在训练集中的次数；$p(\rho)$ 为不管残基类型，局域密度 ρ 被观察到的次数；T_e为有效温度；密度 ρ 为以残基某个特定取向为中心的“自由体积”内β碳原子的密度。自由体积即未被侧链排除的平均体积

$$\rho(\alpha)=\frac{n_\beta}{\frac{4}{3}\pi R^3-\langle V_{access}(\alpha)\rangle}$$

式中，n_β为从侧链质心起在距离 R 内（比如 8Å）的β碳原子数；$\langle V_{access}(\alpha)\rangle$ 为残基 α 的平均排斥体积，是对 α 的所有旋转异构态整体平均值的计算。我们注意局部密度依赖于残基的旋转异构态，所以 $\varepsilon_{env}\{\alpha, \rho[r_k(\alpha)]\}\equiv\varepsilon_{env}\{\alpha, r_k(\alpha)\}$。这种基于β碳原子密度的位能与其他的氨基酸疏水标度相关得很好[59]。对序列概率计算，E_{env} 限制取值为一个具有同样结构的已知序列的值（如果有一个已知值的话），或者具有同

样大小或链长的蛋白质代表值。

1.2.3.3 参考能量

在蛋白质设计中，我们寻求优化特定序列在目标结构中相对于未折叠态系综的能量。为了处理未折叠态，对每一个氨基酸引进一个参考能 γ_{ref}（α）到 E_c中，以模仿失活态的效果[51,65]。这个能量计算为每个氨基酸 α 在 *N*-乙酰基-α-*N'*-甲胺氨基酸形态下的“自由能”，并平均多种主链结构。这是对伸展的未折叠状态的平均值的粗略近似。参考能包含对可能的旋转异构态和可能的主链构象求和。主链构象的 Φ 和 ϕ 每隔 10°取值一次。每个残基的参考能可以用下式估计：

$$\gamma_{ref}(\alpha,\ \beta_{ref}) = -\beta_{ref}^{-1}\ln\left[\frac{z_{ref}(\alpha,\ \beta_{ref})}{z_{ref}(G,\ \beta_{ref})}\right]$$

$$z_{ref}(\alpha,\ \beta_{ref}) = \sum_{\phi,\Phi,k}\exp\{-\beta_{ref}\varepsilon_{ref}[\phi,\ \Phi,\ r_k(\alpha)]\}$$

式中，ε_{ref}为用分子位能确定的 *N*-乙酰基-α-*N'*-甲胺形态的氨基酸 α 在特定的构象的构象能。此处$\beta_{ref}=1/(k_B T)$，其中 k_B是 Boltzmann 常数，而 *T* 是适合于侧链和主链构象采样的温度（如 T＝300K）。这里，始终使用依赖于主链的旋转异构体数据库[60]。参考能是相对于没有侧链的甘氨酸（G）测量。对主链的能量限制包括原子间相互作用并取下列形式

$$\begin{aligned}E_c \approx E_c = &\sum_{i,\alpha,k}\{\varepsilon_i[\alpha,\ r_k(\alpha)] - \gamma_{ref}[\alpha,\ \beta_{ref}]\ w_i[\alpha,\ r_k(\alpha)]\} \\ &+ \sum_{\substack{i,\ j>i \\ \alpha,\ \alpha' \\ k,\ k'}}\varepsilon_{i,j}[\alpha,\ r_k(\alpha);\ \alpha',\ r_{k'}(\alpha')]\ w_i[\alpha,\ r_k(\alpha)]\ w_j[\alpha',\ r_k'(\alpha')]\end{aligned}$$

1.2.3.4 旋转构象和（氨基酸）身份概率

此理论极大化了构象熵 S_c，从而得到特定氨基酸以侧链构象 *k* 出现在位点 *i* 的概率 $w_i[\alpha,\ r(\alpha)]$。氨基酸概率 $w_i(\alpha)$ 可用

$$w_i(\alpha) = \sum_k w_i[\alpha,\ r_k(\alpha)]$$

确定。

类比于统计热力学，由于限制构象能而出现的拉格朗日乘子 β_c，可以认为是等效的倒数温度，即 $1/\beta_c = T_c$。对应的热容量 C_v定义为

$$C_v = \frac{\partial E_c}{\partial T_c} = \beta_c^2\sum_i(\langle\varepsilon_i^2\rangle - \langle\varepsilon_i\rangle^2)$$

$$\langle\varepsilon_i\rangle = \sum_{\alpha,k}\varepsilon_i^{loc}[\alpha,\ r_k(\alpha)]\ w_i[\alpha,\ r_k(\alpha)]$$

$$\varepsilon_i^{loc}[\alpha,r_k(\alpha)] = \varepsilon_i[\alpha,r_k(\alpha)] - \gamma_{ref}(\alpha,\beta_c) + \sum_{j,\alpha',k'}\varepsilon_{i,j}[\alpha,r_k(\alpha);\alpha',r_{k'}(\alpha')]w_j[\alpha',r_{k'}(\alpha')]$$

式中，$\langle\varepsilon_i\rangle$ 为局域平均场能量，它表示围绕特定氨基酸 *i* 的平均局域场；当限制条件的值，例如总能量变化的时候在计算中可被调控时，等效热熔 C_v 便可被用来对序列/旋转异构涨落进行定量。

对特定蛋白——SH3 结构域应用这个理论作为例子。随着有效温度 T_c的降低（即 E_c降低；图 1.1），构象熵也降低。在高能量下（高 T_c，低 β_c），会产生许多残基间不利的相互作用（高能量），以及序列/旋转异构态在每一位点的宽泛分布。一般来说，在每个位点上可能的氨基酸和旋转异构态的数目，随能量的降低而降低。如图 1.1 所示，

C_v在 T_c（$=1/\beta_c$）$=10$mol/kcal 通过一个峰值之后，大约在 2mol/kcal 达到低谷。在这个点内部的残基类型和构象态相对地较为确定[59]；而表面残基，尽管以亲水为主，仍有大量概率可比的旋转异构态。这是与数据库中蛋白质表面暴露残基的构象可变性相一致的。于是，对于确定在哪一个"有效温度"下检验概率，热容量是有帮助的。另外，与序列比对得到的剖面（如用同源性导出的蛋白质二级结构数据库；参考文献［66］）可直接比较得到很符合的结果，特别是在掩埋区（图 1.2）。

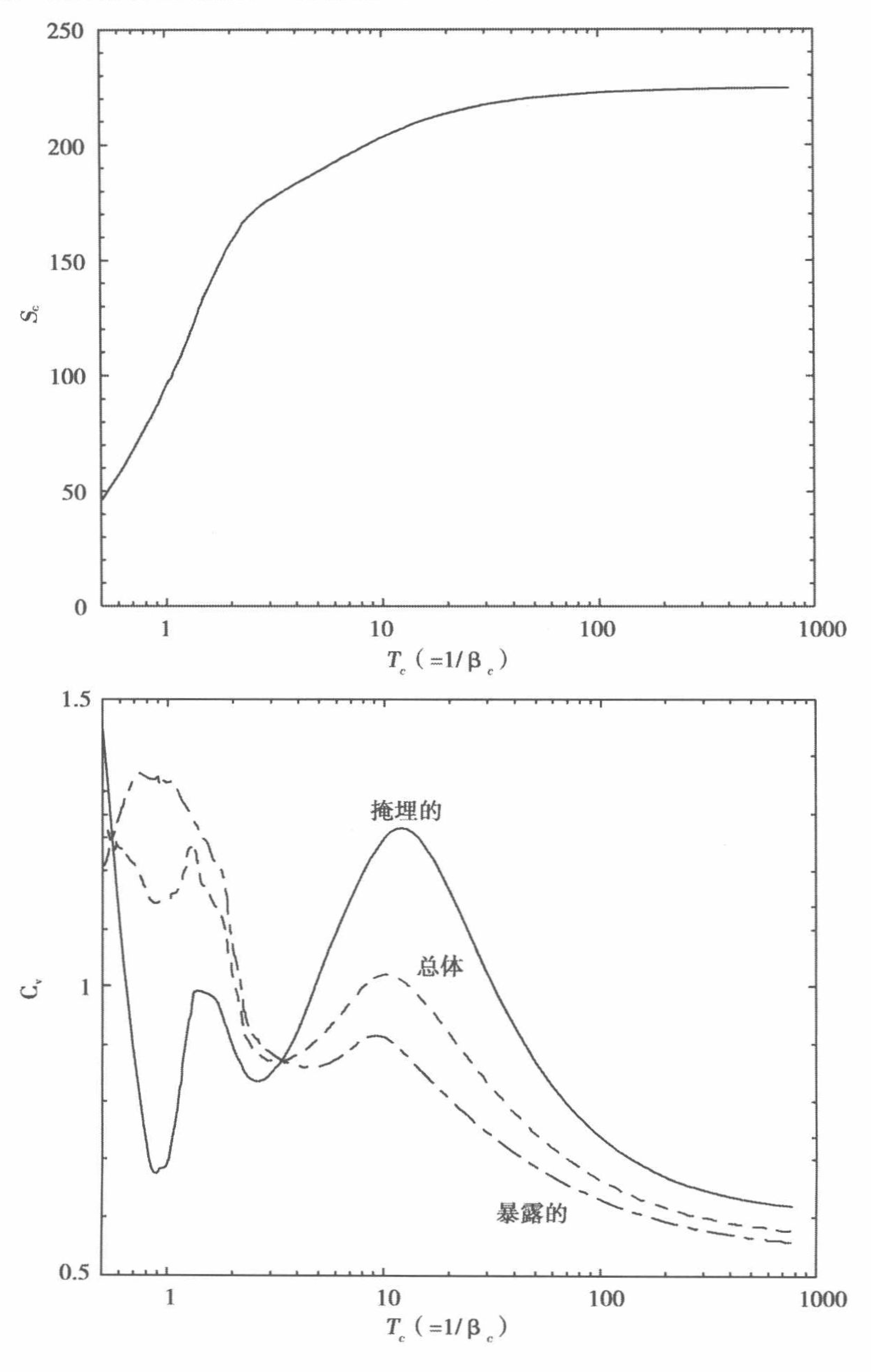


图 1.1　SH3 结构域的序列构象熵 S_c与有效温度 T_c的关系图（上）；对所有残基（包括掩埋的和暴露的）而言，每残基有效热容量与有效温度 T_c的关系见下图。温度由分子位能确定的任意单位给出，此处是 mol/kcal。

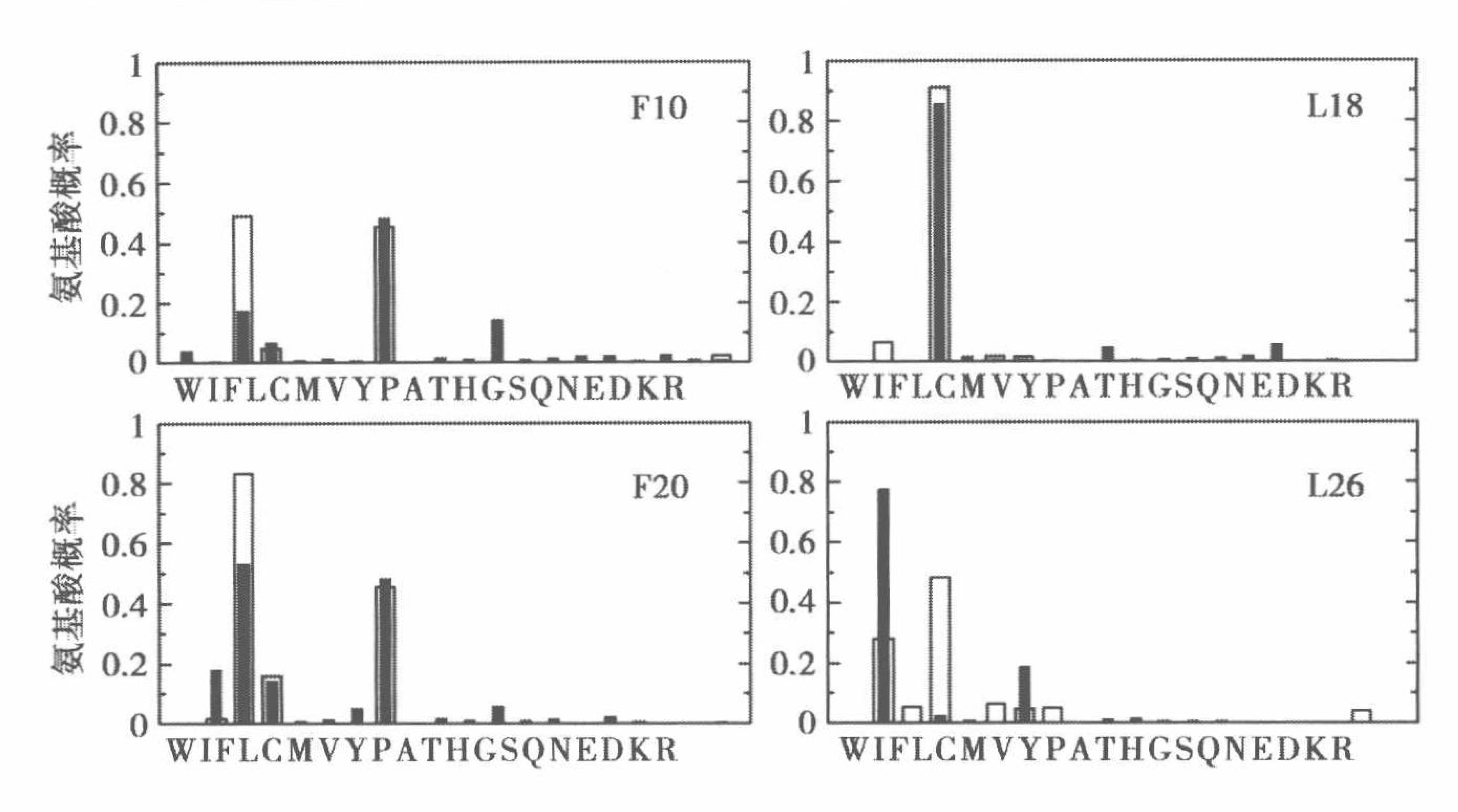


图 1.2 SH3 结构域（PDB：1CKA）的氨基酸概率。(▮) 计算剖面；(□) 基于序列比对的剖面。F10、L18、F20 和 L26 位于疏水核心，溶液可及面积分数小于 20%。

1.2.4 用于蛋白质剖面的基因库

在 1.1.3 小节，我们讨论了怎样应用特定结构的序列剖面。如果公用序列或定向搜索序列鉴定出了一个序列，它可以用多肽合成或合成编码该序列的基因然后再通过表达来直接地实现。大蛋白常用表达的方法来实现。如果概率性的序列信息被用来建立组合库（见注 4），需要有方法把蛋白质剖面转录为部分随机的基因序列库。核苷酸的非均匀分布，对编码偏向于特定氨基酸的多肽序列是必需的。在一组部分随机基因的每个位置的伪独立核苷酸概率可以用计算来确定，使得基因库编码的蛋白质库最好地重现期望的氨基酸剖面。然后，计算出的基因库可用于标准的 DNA 合成中。

令 $P_1(n_1)$，$P_2(n_2)$，$P_3(n_3)$（n_i＝A、T、G 和 C）分别为 4 个核苷酸中的每一个在一个编码的第一、第二和第三个位置的概率。如果认为这些概率是独立的，氨基酸 α 看来会被三联码 $n_1n_2n_3$ 编码的概率是 $P(\alpha \mid n_1, n_2, n_3) = P_1(n_1)P_2(n_2)P_3(n_3)\delta(\alpha \mid n_1, n_2, n_3)$，其中仅当 $n_1n_2n_3$ 是氨基酸 α 的编码时，$\delta(\alpha \mid n_1, n_2, n_3)=1$，否则为零。如果氨基酸 α 的各种编码是等概率的（没有编码偏向），氨基酸 α 的概率为对应于这个氨基酸的各种编码的概率之和：

$$P_{calc}(\alpha) = \sum_{n_1,n_2,n_3} P_1(n_1)P_2(n_2)P_3(n_3)\delta(\alpha \mid n_1, n_2, n_3)$$

目标函数使期望的氨基酸概率分布与一给定的核苷酸概率分布编码的氨基酸概率分布之差定量化[67,68]。为了找到最好地再现期望的氨基酸频度而且避免终止码出现的核苷酸概率，Wang 等提出了一种新的目标函数[69]。此目标函数包含两项，一项为 χ^2 函数——它定量化期望和计算的氨基酸概率的绝对偏差，另一项为相对熵。这样的相对熵广泛地用于定量化两个概率分布间的“距离”，并且是判断一个分布的信息没有被包含在另一个分布中的强硬指标[50]：

$$H=\sum_{\alpha=1}^{21}\left\{P_{calc}(\alpha)\ln\frac{P_{calc}(\alpha)+\varepsilon}{P_{des}(\alpha)+\varepsilon}+0.5\left[P_{des}(\alpha)-P_{calc}(\alpha)\right]^2\right\}$$

式中，ε 是一个任意小的常数（$\varepsilon=10^{-6}$），在 $P_{des}(\alpha)$ 为零的情况下避免数值计算的不

稳定。终止码处理为一个“等效的氨基酸”。在通常的对核苷酸概率的限制条件 $0 \leqslant P_i(n_i) \leqslant 1$ 和 $\sum_{n_i} P_i(n_i) = 1$ 下，优化（极小化）目标函数。这可以用拉格朗日乘子法来完成，或者用限制性极小化的计算程序包[69]。对特殊生物体或表达系统的编码优化也可以包含在这种类型的目标函数中[69]。

图 1.3 说明在特定蛋白质特定氨基酸位点，此处是 SH3 结构域的位点 54 的核苷酸设计。显示出来的是氨基酸的期望的频度（图 1.3 上的空心带）和用于计算的核苷酸概率编码的氨基酸频度（图 1.3 上的实心带)。在这个例子中两者符合得很好。因为氨基酸编码部分简并的原因，在很多情况下，期望和计算的概率分布不能精确地符合。这一计算方法提供了出色的完整序列（不包含终止码)：对于测试蛋白质的 50～60 个残基，若是经历过随机化的选择，其完整序列的产出会是 96%以上。当基因的大部分或全部经历了组合替换，高的产出会显得尤为突出。

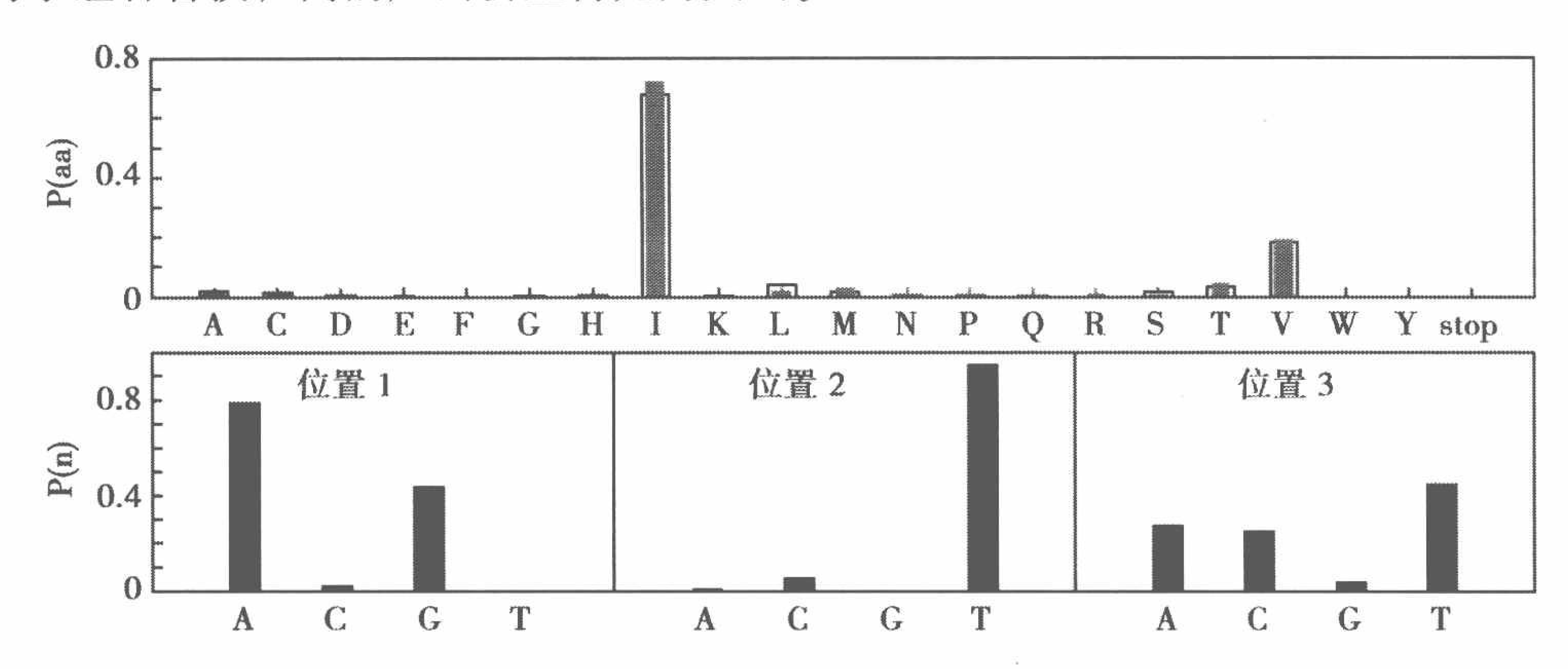


图 1.3　SH3 结构域（PDB：1CKA）位点 54 上的氨基酸（上）和核苷酸（下）概率分布。对氨基酸概率，期望的概率分布用空心带显示，计算的基因库编码的概率分布用实心带显示。具有图下给定的寡核苷酸频度的核苷酸库编码图上显示的位点特异的氨基酸概率。

1.3　注

1. 自洽的方法对整体优化不是最好的[16]。对蛋白质设计，这通常是因为在类平均场理论中处理残基身份和构象态之间相关的近似方式，随机取样和排除法常提供较好的结果。

2. 在设计序列时，主链柔性必须考虑。在天然蛋白质中，主链的微小调整会导致突变[70,71]。但是，大多数计算方法把主链处理为刚性的，因为包括进柔性要消耗大量计算资源。统计理论的概率特性提示其预测剖面对主链选择较不敏感。当能量（或有效温度）在热容量 C_v 的峰值[59]之上，从蛋白 L 的 21 个稍有差别的主链结构得到的氨基酸剖面都相似。对微小的结构涨落均方差为 1Å 紧密相关的序列特性会出现在 E_c 值达到或高于这个热容量峰值时。

3. 在完成熵受限制的优化时，我们发现利用这些方程的寻根法与限制性极小化算法表现得同样好。

4. 氨基酸的计算概率也可以用来计算对突变的容忍度。这样的突变在体外的定向

蛋白进化实验[72]中在具有宽泛的氨基酸分布位点（在这些位点上许多氨基酸具有与最可能的氨基酸可比较的概率）最容易积累。

参考文献

1. Go, N. (1983) Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* 12, 183-210.
2. Shea, J. E. and Brooks, C. L. 3rd. (2001) From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* 52, 499-535.
3. Brooks, C. L. 3rd. (2002) Protein and peptide folding explored with molecular simulations. *Acc. Chem. Res.* 35, 447-454.
4. Kraemer-Pecore, C. M., Wollacott, A. M., and Desjarlais, J. R. (2001) Computational protein design. *Curr. Opin. Chem. Biol.* 5, 690-695.
5. Bryson, J. W., Betz, S. F., Lu, H. S., et al. (1995) Protein design: a hierarchic approach. *Science* 270, 935-941.
6. Dunbrack, R. (2002) Rotamer libraries. *Curr. Opin. Struct. Biol.* 12, 431-440.
7. Shakhnovich, E. I. and Gutin, A. M. (1993) A new approach to the design of stable proteins. *Protein Eng.* 6, 793-800.
8. Jones, D. T. (1994) *De novo* protein design using pairwise potentials and a genetic algorithm. *Protein Sci.* 3, 567-574.
9. Hellinga, H. W. and Richards, F. M. (1994) Optimal sequence selection in proteins of known structure by simulated evolution. *Proc. Natl. Acad. Sci. USA* 91, 5803-5807.
10. Desjarlais, J. R. and Handel, T. M. (1995) *De novo* design of the hydrophobic cores of proteins. *Protein Sci.* 4, 2006-2018.
11. Johnson, E. C., Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1999) Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Struct. Fold. Des.* 7, 967-976.
12. Jiang, X., Farid, H., Pistor, E., and Farid, R. S. (2000) A new approach to the design of uniquely folded thermally stable proteins. *Protein Sci.* 9, 403-416.
13. Jiang, X., Bishop, E. J., and Farid, R. S. (1997) A *de novo* designed protein with properties that characterize natural hyperthermophilic proteins. *J. Am. Chem. Soc.* 119, 838, 839.
14. Bryson, J. W., Desjarlais, J. R., Handel, T. M., and DeGrado, W. F. (1998) From coiled coils to small globular proteins: design of a native-like three-helix bundle. *Protein Sci.* 7, 1404-1414.
15. Walsh, S. T. R., Cheng, H., Bryson, J. W., Roder, H., and DeGrado, W. F. (1999) Solution structure and dynamics of a *de novo* designed three-helix bundle protein. *Proc. Natl. Acad. Sci. USA* 96, 5486-5491.
16. Voigt, C. A., Gordon, D. B., and Mayo, S. L. (2000) Trading accuracy for speed: a quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.* 299, 789-803.
17. Gordon, D. B. and Mayo, S. L. (1998) Radical performance enhancements for combinatorial optimization algorithms based on the dead-end elimination theorem. *J. Comput. Chem.* 19, 1505-1514.
18. Gordon, D. B. and Mayo, S. L. (1999) Branch-and terminate: a combinatorial optimization algorithm for protein design. *Struct. Fold. Des.* 7, 1089-1098.
19. Pierce, N. A., Spriet, J. A., Desmet, J., and Mayo, S. L. (2000) Conformational splitting: a more powerful criterion for dead-end elimination. *J. Comput. Chem.* 21, 999-1009.
20. Looger, L. L. and Hellinga, H. W. (2001) Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J. Mol. Biol.* 307, 429-445.
21. Dahiyat, B. I. and Mayo, S. L. (1997) *De novo* protein design: fully automated sequence selection. *Science* 278, 82-87.
22. Marshall, S. A. and Mayo, S. L. (2001) Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol.* 305, 619-631.

23. Malakauskas, S. M. and Mayo, S. L. (1998) Design, structure, and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* 5, 470-475.

24. Strop, P. and Mayo, S. L. (1999) Rubredoxin variant folds without iron. *J. Am. Chem. Soc.* 121, 2341-2345.

25. Shimaoka, M., Shifman, J. M., Jing, H., Takagi, L., Mayo, S. L., and Springer, T. A. (2000) Computational design of an integrin i domain stabilized in the open high affinity conformation. *Nat. Struct. Biol.* 7, 674-678.

26. Benson, D. E., Wisz, M. S., Liu, W., and Hellinga, H. W. (1998) Construction of a novel redox protein by rational design: conversion of a disulfide bridge into a mononuclear iron-sulfur center. *Biochemistry* 37, 7070-7076.

27. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F., and Lombardi, A. (1999) *De novo* design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* 68, 779-819.

28. Bolon, D. N. and Mayo, S. L. (2001) Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA* 98, 14274-14279.

29. Street, A. G. and Mayo, S. L. (1999) Computational protein design. *Struct. Fold. Des.* 7, R 105-R 109.

30. Saven, J. G. (2001) Designing protein energy landscapes. *Chem. Rev.* 101, 3113-3130.

31. Gromiha, M. M., Uedaira, H., An, J., Selvaraj, S., Prabakaran, P., and Sarai, A. (2002) Protherm, thermodynamic database for proteins and mutants: developments in version 3.0. *Nucleic Acids Res.* 30, 301, 302.

32. Calhoun, J. R., Kono, H., Lahr, S., Wang, W., DeGrado, W. F., and Saven, J. G. (2003) Computational design and characterization of a monomeric helical dinuclear metalloprotein. *J. Mol. Biol.* 334, 1101-1115.

33. Slovic, A. M., Kono, H., Lear, J. D., Saven, J. G., and DeGrado, W. F. (2004) Computational design of water-soluble analogues of the potassium channel kcsa. *Proc. Natl. Acad. Sci. USA* 101, 1828-1833.

34. Zou, J. and Saven, J. G. (2003) Using self-consistent fields to bias monte carlo methods with applications to designing and sampling protein sequences. *J. Chem. Phys.* 118, 3843-3854.

35. Park, S., Kono, H., Wang, W., Boder, E. T., and Saven, J. G. (2005) Progress in the development and application of computational methods for probabilistic protein design. *Comp. Chem. Eng.* 24, 407-421.

36. Keefe, A. D. and Szostak, J. W. (2001) Functional proteins from a randomsequence library. *Nature* 410, 715-718.

37. Rojas, N. R. L., Kamtekar, S., Simons, C. T., et al. (1997) *De novo* heme proteins from designed combinatorial libraries. *Protein Sci.* 6, 2512-2524.

38. Roy, S., Ratnaswamy, G., Boice, J. A., Fairman, R., McLendon, G., and Hecht, M. H. (1997) A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. *J. Am. Chem. Soc.* 119, 5302-5306.

39. Roy, S., Helmer, K. J., and Hecht, M. H. (1997) Detecting native-like properties in combinatorial libraries of *de novo* proteins. *Fold. Des.* 2, 89-92.

40. Finucane, M. D., Tuna, M., Lees, J. H., and Woolfson, D. N. (1999) Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry* 38, 11, 604, 11, 612.

41. Xu, G. F., Wang, W. X., Groves, J. T., and Hecht, M. H. (2001) Self-assembled monolayers from a designed combinatorial library of *de novo* beta-sheet proteins. *Proc. Natl. Acad. Sci. USA* 98, 3652-3657.

42. Case, M. A. and McLendon, G. L. (2000) A virtual library approach to investigate protein folding and internal packing. *J. Am. Chem. Soc.* 122, 8089, 8090.

43. Arndt, K. M., Pelletier, J. N., Müller, K. M., Alber, T., Michnick, S. W., and Plückthun, A. (2000) A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. *J. Mol. Biol.* 295, 627-639.

44. Zhao, H. M. and Arnold, F. H. (1997) Combinatorial protein design: strategies for screening protein libraries. *Curr. Opin. Struct. Biol.* 7, 480--485.

45. Giver, L. and Arnold, F. H. (1998) Combinatorial protein design by in vitro recombination. *Curr. Opin. Chem. Biol.* 2, 335-338.

46. Hoess, R. H. (2001) Protein design and phage display. *Chem. Rev.* 101, 3205-3218.

47. Moffet, D. A. and Hecht, M. H. (2001) *De novo* proteins from combinatorial libraries. *Chem. Rev.* 101, 3191-3203.

48. Holm, L. and Sander, C. (1998) Touring protein fold space with dali/fssp. *Nucleic Acids Res.* 26, 316-319.

49. Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) Assessment of protein models with 3-dimensional profiles. *Nature* 356, 83-85.

50. Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge, UK.

51. Raha, K., Wollacott, A. M., Italia, M. J., and Desjarlais, J. R. (2000) Prediction of amino acid sequence from structure. *Protein Sci.* 9, 1106-1119.

52. Kuhlman, B. and Baker, D. (2000) Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* 97, 10383-10388.

53. Koehl, P. L. M. (1999) *De novo* protein design. I. In search of stability and specificity. *J. Mol. Biol.* 239, 1161-1181.

54. Koehl, P. L. M. (1999) *De novo* protein design. II. Plasticity in sequence space. *J. Mol. Biol.* 293, 1183-1193.

55. Kraemer-Pecore, C. M., Lecomte, J. T., and Desjarlais, J. R. (2003) A *de novo* redesign of the ww domain. *Protein Sci.* 12, 2194-2205.

56. Larson, S. M., England, J. L., Desjarlais, J. R., and Pande, V. S. (2002) Thoroughly sampling sequence space: large-scale protein design of structural ensembles. *Protein Sci.* 11, 2804-2813.

57. Hayes, R. J., Bentzien, J., Ary, M. L., Hwang, M. Y., Jacinto, J. M., Vielmetter, J., Kundu, A., and Dahiyat, B. I. (2002) Combining computational and experimental screening for rapid optimization of protein properties. *Proc. Natl. Acad. Sci. USA* 99, 15926-15931.

58. Zou, J. and Saven, J. G. (2000) Statistical theory of combinatorial libraries of folding proteins: energetic discrimination of a target structure. *J. Mol. Biol.* 296, 281-294.

59. Kono, H. and Saven, J. G. (2001) Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *J. Mol. Biol.* 306, 607-628.

60. Dunbrack, R. and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain retainer preferences. *Protein Sci.* 6, 1661-1681.

61. McQuarrie, D. A. (1976) *Statistical mechanics.* Harper and Row, New York.

62. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) *Numerical recipes.* 2nd ed., Cambridge University Press, Cambridge, UK.

63. Weiner, S. J., Kollman, P. A., Case, D. A., et al. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106, 765-784.

64. Kono, H. and Doi, J. (1996) A new method for side-chain conformation prediction using a hopfield network and reproduced rotamers. *J. Comput. Chem.* 17, 1667-1683.

65. Wernisch, L., Hery, S., and Wodak, S. J. (2000) Automatic protein design with all atom force-fields by exact and heuristic optimization. *J. Mol. Biol.* 301, 713-736.

66. Sander, C. and Schneider, R. (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 9, 56-68.

67. Jensen, L. J., Andersen, K. V., Svendsen, A., and Kretzschmar, T. (1998) Scoring functions for computational algorithms applicable to the design of spiked oligonucleotides. *Nucleic Acids Res.* 26, 697-702.

68. Wolf, E. and Kim, P S. (1999) Combinatorial codons: a computer program to approximate amino acid probabilities with biased nucleotide usage. *Protein Sci.* 8, 680-688.

69. Wang, W. and Saven, J. G. (2002) Designing gene libraries from protein profiles for combinatorial protein experiments. *Nucleic Acids Res.* 30, e 120.

70. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic

effect. *Science* 255, 178-183.

71. Axe, D. D., Foster, N. W., and Fersht, A. R. (1996) Active barnase variants with completely random hydrophobic cores. *Proc. Natl. Acad. Sci. USA* 93, 5590-5594.

72. Voigt, C. A., Mayo, S. L., Arnold, F. H., and Wang, Z. G. (2001) Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. USA* 98, 3778-3783.

2 大肠杆菌中非天然氨基酸的整体掺入

Jamie M. Bacher and Andrew D. Ellington

概要 氨基酸类似物的掺入越来越有用。非天然氨基酸的定点掺入，使得运用化学生物学对特定蛋白质的研究和应用成为可能。但是，非天然氨基酸的整体掺入也检验着蛋白质组和基因编码假定。例如，有机体对非天然氨基酸的适应可能会导致新的基因编码。为了理解和定量化这样的掺入引起的变化，需理解微生物学和蛋白质组对非天然氨基酸掺入的反应。此文描述了在蛋白质组范围鉴定这些掺入效应的规程。

关键词 非天然氨基酸；基因编码的模糊性；氨基酸的误掺入；氨酰化错误；氨基酸类似物；基因编码进化

2.1 介　　绍

包含非天然氨基酸的蛋白质的过量表达已引起越来越多科学家的关注。作为一个例子，非天然氨基酸的定点掺入，使关于蛋白质结构和功能的化学假定的特异检测成为可能。类似地，具有新化学性质的非天然氨基酸的定点掺入，可能使蛋白质产生新的功能，如与酮基取代的氨基酸交联[1]。对整个蛋白质的整体非天然氨基酸掺入，也可能导致新的物理或功能性质。例如，用硒代甲硫氨酸那样的重原子类似物代替甲硫氨酸，已成为 X 射线晶体学中得到差值图（进而得到位相，解析蛋白质晶体结构——译者注）的有用工具[2]。用非天然氨基酸整体地扰动生物体的蛋白质组，已被用来在实验中探测基因编码的进化。细菌和噬菌体已适应于非天然氨基酸的掺入，并且为这些进化转变所需要的突变的数量和类型也已经过检验[3~6]。

虽然有大量文献涉及细菌在非天然氨基酸环境中的生长（如用色氨酸类似物，见参考文献［7～11］)，但其中大部分只是叙述技术，而没有考虑工艺规程的改变怎样影响实验结果。为此，我们在这里提供一些更详尽的方法。

2.2 材　　料

（1）大肠杆菌菌株 C600 Δ *trp*E（*thi*-1 *thr*-1 *leu*B6 *lac*Y1 *ton*A21 *sup*E44 mcrA Δ *trp*E）和衍生菌株 C600p（C600 Δ *trp*E＋pUC18)、C600pGSR（C600 Δ *trp*E＋pG-SR)、C600F（C600 Δ *trp*E F′KanR）和 C600F（DE3)（C600FλDE3 溶源体)。用于转化的菌株有 DH5αF′和 TOP10（Invitrogen, Carlsbad, CA)。

（2）Luria-Bartani 培养基（每升含：10g 胰蛋白胨、5g 酵母抽提物、10g 氯化钠和 1.5%菌用琼脂糖用于培养板）和基本培养基 M9（5×原液，每升含：30g 磷酸氢二钠、

15g 磷酸二氢钾、5g 氯化铵、2.5g 氯化钠和 1.5%菌用琼脂糖用于培养板），补充以 20μg/ml 氨基酸（见注 1）和 0.0005%硫胺。富培养基和基本培养基都如标示的那样补充以抗生素：50μg/ml 氨苄青霉素（Ap）或卡那霉素（Kn）。

（3）色氨酸类似物：4-、5-和 6-氟色氨酸（fW），Sigma（St. Louis，MO）。

（4）质粒：pGEX-KG[12]和 pET100/D-topo（Invitrogen），用于蛋白质高表达。用于基因的聚合酶链反应（PCR）扩增：pGFPuv（Clontech，La Jolla，CA）或另外来源的编码 GFPuv 基因和质粒来源的 Kn 激酶基因（如 p182Sfi-，Kan，K. A. Marshall 和 A. D. Ellington，未发表结果）。

（5）Vent 和 *Taq* DNA 聚合酶，限制性内切核酸酶，DNA 酶，T4 激酶和 T4 DNA 连接酶。

（6）寡核苷酸引物和 dNTP（Invitrogen）。

（7）细菌蛋白抽提物试剂（B-PER）和 B-PER II（Pierce，Beverly，MA）。

（8）100mmol 溶于水的异丙基-β-D-硫代半乳糖苷（IPTG）作为储备液。

（9）Microcon 浓缩器，拦截的相对分子质量为 10 000（Microcon，Rockford，IL）。

（10）谷胱甘肽-琼脂糖糖球珠（Amersham Biosciences，Piscataway，NJ）。

（11）Ni-次氮基三乙酸（NTA）树脂（Novagen）和蛋白质纯化柱（Bio-Rad，Hercules，CA）。

（12）Centri-Sep 尺寸排斥柱（Princeton Separation，Adelphia，NJ）。

（13）L-甲苯磺酰氨-2-苯乙基氯甲基酮处理过的胰蛋白酶（Pierce，Beverly，MA）。

（14）磷酸缓冲液（PBS）：10×储备液，每升含：80g 氯化钠、2g 氯化钾、11.5g 磷酸氢二钠水合物（$Na_2HPO_4 \cdot 7H_2O$）和 2g 磷酸二氢钾。

（15）1mol/L 氯化镁。

（16）50mmol/L Tris-氯化氢，pH 8.0 和 5mmol/L 还原谷胱甘肽。

（17）Ni-NTA 纯化用的缓冲液：

a. 结合缓冲液，8×储备液：160mmol/L Tris-氯化氢，pH 7.9，4mol/L 氯化钠和 40mmol/L 咪唑。

b. 清洗缓冲液，8×储备液：160mmol/L Tris-氯化氢，pH 7.9，4mol/L 氯化钠和 480mmol/L 咪唑。

c. 洗脱缓冲液，8×储备液：160mmol/L Tris-氯化氢，pH 7.9，4mol/L 氯化钠和 2mol/L 咪唑。

（18）高效液相层析（HPLC）分析用的缓冲液：

a. 缓冲液 A：50mmol/L NH_4OAc，pH 5.0。

b. 缓冲液 B：50mmol/L NH_4OAc，pH 5.0 和 50% MeOH。

c. 缓冲液 C：0.1mol/L 磷酸二氢钠，pH 2.5。

d. 缓冲液 D：0.1mol/L 磷酸二氢钠，pH 2.5 和 50% MeOH。

（19）琼脂糖 DNA 凝胶设备和十二烷基硫酸钠（SDS）凝胶电泳（PAGE）设备。

（20）HPLC、HPLC-电喷电离（ESI）和质谱设备。

（21）微板阅读器。

2.3　方　　法

下述方法包括：

（1）大肠杆菌在色氨酸类似物中的生长。

（2）含 fW 取代氨基酸的蛋白质的表达和纯化。

（3）对非天然氨基酸进行深入水平的分析。

2.3.1　大肠杆菌在非天然氨基酸（培养基）上的生长

色氨酸类似物掺入大肠杆菌是直截了当的。因为色氨酰转移 RNA 合成酶没有编辑结构域，天然氨基酸与类似物的区分只是在结构上。枯草杆菌色氨酰转移 RNA 合成酶可以相对更有效地结合（装载）荧光化的色氨酸类似物；4fW 的结合比 W 低 6 倍，而 5fW 的结合比 W 低 74 倍[13]。与此类似，已经知道色氨酸类似物进入细胞，并在类似物的毒性起作用之前支持最少几代的生长。为了使色氨酸类似物有效掺入，使用带有防止生物合成的突变细菌菌株（见注 2；参考文献［4～6］，［9］和［11］）。在本节介绍的例子中，使用了大肠杆菌菌株 C600 Δ *trp*E 及其衍生菌株。基本培养基中补充了苏氨酸、亮氨酸、硫胺以及色氨酸类似物，或某些比例的非天然对天然氨基酸。习惯上，培养基是用补充成分命名的。例如，M9B1TL95% 4fW＋Ap 补充了维生素 B1（硫胺）、苏氨酸、亮氨酸、19∶1 的 4fW∶W 和 Ap。

非天然氨基酸掺入对细菌生长能力的影响可通过分析生长曲线来确定。ELX808 微板阅读器[14]或 Bioscreen C[5]这样的仪器可以并行地获得多条微板格式的生长曲线。在用 Bioscreen C 的测试中，C600p 在基本培养基中的过夜培养液稀释 100 倍后，接种到每种 1.5ml 的测试培养基中，以测试对生长的抑制。将这些 1.5ml 的培养基按 350μl 分到 3 个小坑中。于 37℃不停地摇动，直到所有培养物均进入稳定生长期。把生长曲线的指数部分拟合到方程：

$$N_{(t)}=N_{(0)}\times e^{(-rt)} \qquad (2.1)$$

式中，t 为时间；$N_{(t)}$为时刻 t 的菌体数（或光密度）；$N_{(0)}$是初始的菌体数，可以算出菌体固有的生长速率 r。例如，在生长曲线的指数部分拟合到式（2.1）以后，在对 W 的比率在 97% 以下时，4fW 很少引起生长速率的改变，而 5fW 和 6fW 的影响要激烈得多（图 2.1）。因此，作为细菌的常规生长，比率 19∶1 4fW∶W 可以在培养基中使用。如果培养基只提供 4fW 的话，大肠杆菌可以连续几代生长。而且，通过拟合整条曲线（与仅拟合指数期相反）到对数生长方程，可以达到对生长曲线的更完全理解。这个对数生长方程考虑了承载能力：

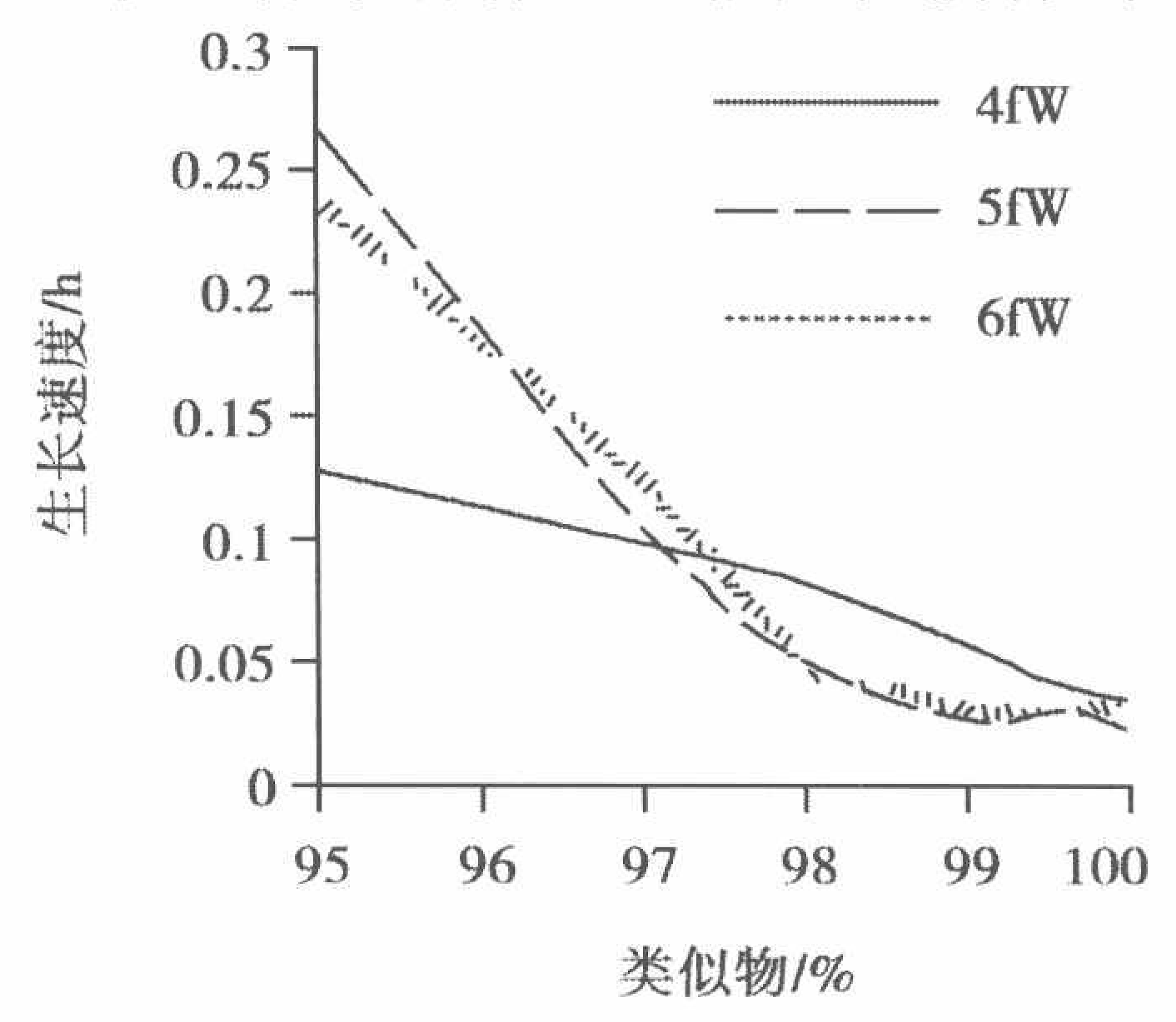


图 2.1　各种比例的类似物与天然色氨酸对大肠杆菌生长的效应。菌株 C600p 生长在各种比例的 4fW，5fW 和 6fW 的条件之下。生长用微板中的光谱仪跟踪，速率用 Eq.1 确定（数据在参考文献［5］的允许下重印，但用这个方程重新分析过）。

$$N_{(t)}=\frac{K}{1+\{[(K/N_{(0)})-1]\times e^{-rt}\}} \quad (2.2)$$

式中，除了培养物的承载能力 K 以外，其他变量与式（2.1）相同。这个方程已用于在氨酰化错误条件下大肠杆菌氨基酸类似物掺入效果的定量化[15]。

2.3.2　掺入 4fW 的蛋白质的表达和纯化

本节描述建立两个表达质粒（见 3.2.1.1 和 3.2.1.2）以及在 W 替换为 4fW 的高置换率条件下蛋白质表达和纯化的步骤（见 3.2.2.1 和 3.2.2.2）。用表达和纯化的语言，在氨基酸类似物上的生长所需要的工艺变动，与只在天然氨基酸上的生长略有关系。最后，这一节将对从已掺入非天然氨基酸的细菌中分离胞内总蛋白的方法给一个概述。

2.3.2.1　表达载体的构建

1）pGSR

由于 C600p 中的 pUC18 质粒 pGEX-KG［一种谷胱甘肽-*S*-转移酶（GST）表达载体］需要 Ap 以外的选择方法，因而，用 *Sma*I 和 *Eco*RI 降解质粒。P182Sfi-Kan 的 Kn 激酶基因由引物 Kan1.39（5′-CGCGGATCCGGCCACCATGGCCAAGCGAACCGGAAT）和 Kan2.39（5′-CCGGAATTCTGAGGCCTGACAGGCCTTAGAAGAACTCGT）用 PCR 扩增。PCR 产物用 *Bsa*BI 和 *Eco*RI 降解并连接到 *Sma*I-和 *Eco*RI 降解后的质粒 pGEX-KG 上[16]，然后转化到 DH5αF′中，再用微量制备分离（QIAgen)。得到的质粒 pGSR 是一种带有 Kn 抗性的 GST 表达载体。

2）pET100GFPuv

高荧光的 GFPuv 的基因从质粒 pGFPuv 用 Vent DNA 聚合酶（NEB）以引物 CF-PA（5′-CACCACGGCCACTGTGGCCATGAGTAAAGGAGAAGAACTT-3′）和 CF-PB（5′-GGCCATCGGGGCCCTATTTTATAGTTCATCCATGCC-3′）通过 PCR 扩增；拓扑异构酶介导的定向克隆需要在正向引物中的 5′-CACC 片段。通过 7.5μl PCR 产物与 1μl 的 10×缓冲液、1μl 的 4mmol dNTP 和 0.5μl 的 *Taq* DNA 聚合酶在 72℃保育 20min，把突出的腺苷残基加在扩增产物上。按制造者的指示，这个反应用来克隆 *GFPuv* 基因到 pET100/D-topo 中。拓扑异构酶反应是用来转化化学上具竞争性的 TOP10 细胞。得到的质粒用微制备（QIAgen）分离。PET100GFPuv 带有 Ap 抗性并且在 T7 核酸（RNA）聚合酶启动子的控制下，需要在携带 λDE3 溶源体的寄主菌株中表达 *GFPuv*。

2.3.2.2　蛋白质表达和纯化

为了实现非天然氨基酸的高水平掺入，初期的生长是在比较宽松的条件下进行的，然后转变到包含诱导子的更具限制性的条件下[5]。这一目的也可以通过提供有限数量的容许氨基酸和过量的非天然氨基酸来实现。当天然氨基酸耗尽时，这些非天然氨基酸就会得到利用（见注 3 和参考文献[17]）。

1）高掺入 4fW 的 GST 的表达和纯化

（1）在 M9B1TL95％4fW＋Kn 中，过夜生长的用 pGSR 转化的并在 Kn 板上选择

的 C600p 的一单菌斑。

(2) 把第一步的初始培养稀释 100 倍接种于 100ml 同样基质的培养液中。把细菌培养到中指数期（mid-log）(在 600nm 处，光密度 0.5)。

(3) 以 5400*g* 离心培养液 20min，并悬浮在补充了 0.3mmol/L IPTG 的 100ml M9B1TL3×100%4fW＋Kn 中，继续培养 16h（见注 4）。

(4) 如步骤 3 将细胞离心，再用 5ml B-PER（Novagen）试剂裂解。

(5) 离心去除不溶解部分之后，加入 10mmol/L 氯化镁（从 1mol 储备液中取）和 5U DNA 酶，常温保育 15min。

(6) 加入 500μl 50%谷胱甘肽琼脂糖凝胶球珠浆到裂解液中，并在常温下用搅拌机混合 2min[18]。

(7) 以最高速度简单离心球珠。加入 5ml PBS 冲洗球珠。再重复冲洗 2 次。

(8) 用 1ml 冰冷的 PBS 做最后的冲洗，再移到微离心管中。

(9) 在室温下用 50mmol/L Tris-氯化氢，pH 8.0，外加 5mmol/L 还原的谷胱甘肽。在室温下旋转 2min 以从球珠上洗脱纯化的 GST，分 3 次进行，每次 0.5ml。

(10) 用 Microcon 浓缩器浓缩这 3 份样品。

(11) 用 SDS-PAGE 测定蛋白质纯度；纯度应高于 95%。

2）掺入 6fW 的 GFPuv 的表达和纯化

(1) 在 200ml 的 M9B1TLW ＋ Kn ＋ Ap 中生长的 C600F（DE3）中加入 pET100GFPuv。

(2) 在中指数期，离心培养并使得到的细胞小团重新悬浮在 100ml 含 1mmol/L IPTG 的 M9B1TL95%6fW＋Kn＋Ap 中再继续生长 3h（见注 4）。

(3) 在终止培养并用 3ml B-PER 裂解细胞小团。以 15 000*g* 离心 30min 以清除不溶物。在上清液中加入 10mmol/L 氯化镁（从 1mol 储备液中取）并加入 3U DNA 酶，在室温保育 10min。

(4) 准备 3ml Ni-NTA 柱，用 15ml 水和 9ml 结合缓冲液冲洗。

(5) 让上清液通过 Ni-NTA 柱。用 30ml 结合缓冲液和 18ml 冲洗缓冲液冲洗，再用 18ml 洗脱缓冲液洗脱。从柱上逐份搜集洗脱液，每份 0.5ml，用 SDS-PAGE 和用紫外光照射分析（含有蛋白质的样品可见荧光）。浓缩含蛋白质的样品如 2.3.2.2 中 1）所述。

2.3.2.3 细胞全蛋白抽提物的纯化

用于 4fW 的掺入：

(1) 在 25ml 恰当的基质中生长 C600p 到饱和；

(2) 离心培养液其变为液滴，并在 200 μl B-PER 中裂解；

(3) 让 50μl 裂解液通过 Centri-Sep 尺寸排斥柱以去掉未掺入的氨基酸。

类似地，用于分析细菌对 6fW 的鉴别：

(1) 在 100μl M9B1TLW＋Kn 培养基或 M9B1TL95%6fW＋Kn 培养基中，使生长的 C600F 达到饱和；

(2) 使细菌聚集成颗粒状并在 200μl B-PER 中裂解；

(3) 半体积的裂解液通过 Centri-Sep 柱［见 2.3.3.1 中 1)］。

2.3.3　非天然氨基酸掺入程度分析

本节讨论了确定非天然氨基酸掺入水平的两种方法。完全水解然后用 HPLC 和质谱分析 [见 2.3.3.1 中 1)]，或者 HPLC 分析之后在不同的条件下再用 HPLC 分析 [见 2.3.3.1 中 2)]，可以用来纯化蛋白质和全细胞蛋白抽提。但是，蛋白质酶解和片段分析（见 2.3.3.2）只能用来纯化蛋白质。

2.3.3.1　纯化蛋白的分析

表 2.1 列举了应用 2.3.3.1 中 1）和 2.3.3.1 中 2）两个小节所述方法的例子。

表 2.1　6fW 之掺入蛋白

分析方法	章节	掺入量/%
全细胞蛋白提取物		
HPLC-ESI	2.3.3.1 中 1)	56.5
HPLC-HPLC	2.3.3.1 中 2)	66.7
纯化的 GFPuv		
HPLC-ESI	2.3.3.1 中 1)	68.6
平均掺入量		64.0

摘自参考文献[4]。

1）含非天然氨基酸蛋白的水解和 HPLC-ESI 分析

蛋白质样品的酸水解可以确定蛋白质的成分。不管是全细胞蛋白抽提物或纯化的蛋白质，都要经过整体平均化。

（1）冻干蛋白样品（如 2.3.2.3 中从 Centri-Sep 柱出来的半量洗脱液，2.3.2.2 中几毫克纯化蛋白）。

（2）将团体物质重悬在 1ml 含 10%巯基乙酸的 5.4mol/L 氯化氢中以在水解中保护色氨酸。

（3）在真空和 110℃下水解过夜。

（4）冻干水解产物，并重悬在 50μl 的水中。

水解产物可用 HPLC-ESI 分析。在这种情况下，当从 HPLC 柱上洗脱时，天然的和非天然的氨基酸的比质量可被测定和跟踪。洗脱质量的相对比率可用曲线下的面积确定。氨基酸的实际摩尔数可用标准曲线确定。

2）水解蛋白样品的 HPLC-HPLC 分析

另一种办法是完成一系列的 HPLC 分析。

（1）把水解试剂注入 C-18 柱并用下列步骤洗脱。

a. 用 94%缓冲液 A 和 6%缓冲液 B 洗脱 20min。

b. 10min 内用梯度法切换到 98%缓冲液 A 和 2%缓冲液 B。

c. 再洗脱 15min。

d. 在最后 5min 将柱子重新平衡到 94%缓冲液 A 和 6%缓冲液 B。

（2）收集对应于标准峰（如 W 和 6fW）的洗脱液并冻干，再重悬在水中。

(3) 把样品重新注入同一个柱中，但缓冲液改为80%缓冲液C和20%缓冲液D。

不同分子的同时洗脱可能发生在一组缓冲液条件下，但不大可能发生在两组缓冲液的条件下。在第二个缓冲条件下的曲线下面积可以认为是纯样品，并可以比较对应于天然的和非天然的峰面积。

在有色氨酸和色氨酸类似物的情形下，280nm处探测，不需要衍生化（见注5）。苯丙氨酸和酪氨酸在纯化的蛋白质中显示小的吸收峰，而全细胞蛋白的抽提物会产生大量的小峰。尽管如此，定量的数据仍可得到。

2.3.3.2 蛋白酶降解以确定非天然氨基酸掺入水平

由于特定多肽片段的质量可以基于蛋白质序列预测，预测片段的质量位移有可能是由于特定非天然氨基酸掺入的结果。因而，分析蛋白酶降解产物就是确定氨基酸类似物掺入的敏感方法[5,19,20]。

(1) 冻干纯化的GST [如5μg，见2.3.2.2中1)] 并重悬在0.1mol/L碳酸氢铵中。

(2) 在37℃用被固定化的L-甲苯磺酰胺-2-苯乙基氯甲基酮处理过的胰蛋白酶（Pierce）降解10h。

(3) 离心以去除胰蛋白酶（见注6）。

(4) 冻干降解物，并在水中重悬至210μmol/L。

(5) 用HPLC-ESI分析降解产物。

可以跟踪特定质量的洗脱剖面，并证明非天然氨基酸在特定片段中的掺入水平（图2.2）。

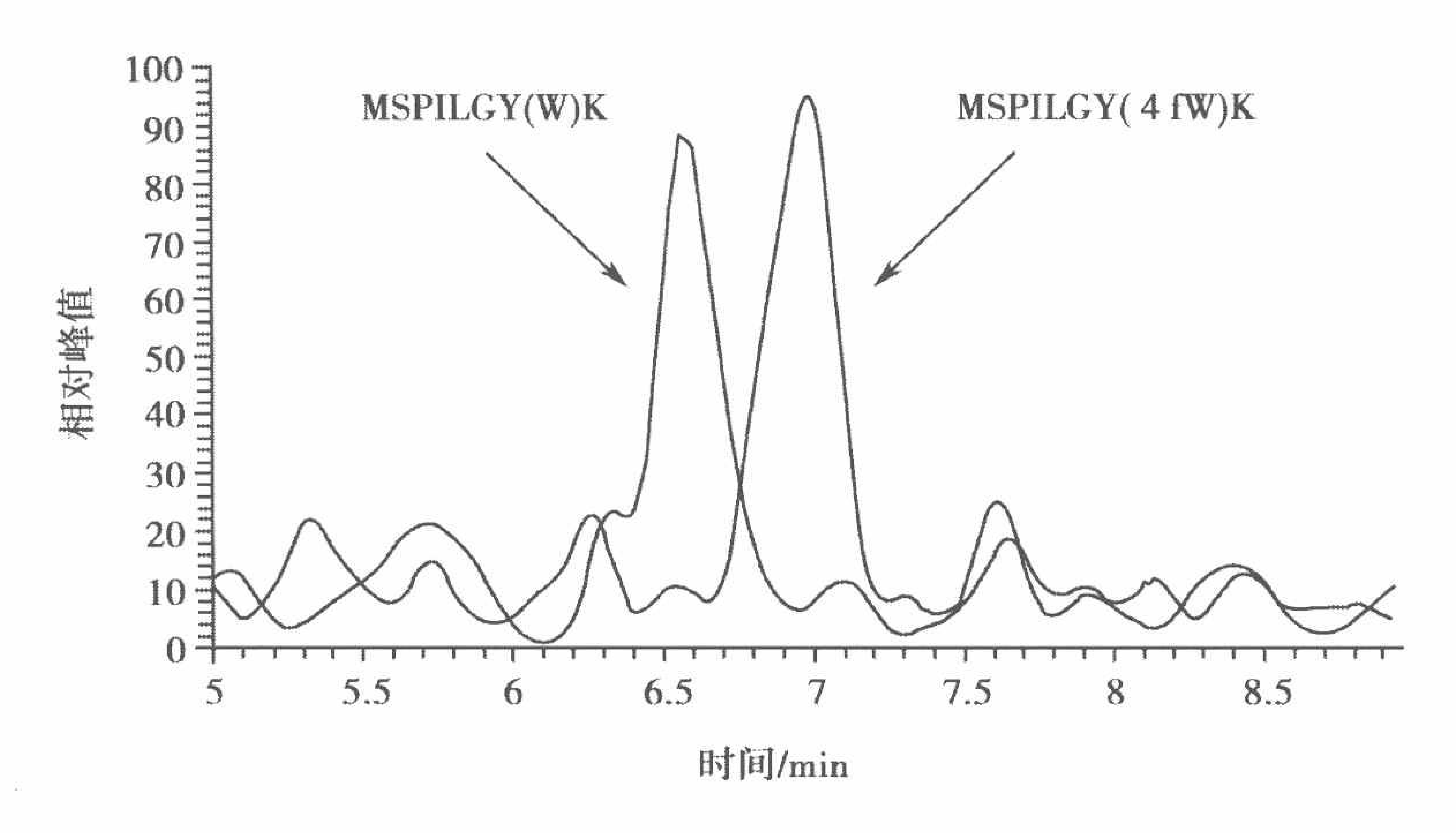


图2.2 4fW和W的相对掺入。从在95%4fW上生长的C600pGSR中提纯的GST用胰蛋白酶降解并用HPLC-ESI分析。多肽MSPILGY（W/4fW）K在分子质量为1094Da和1112Da时产生出比例大致相同的两个峰（经参考文献［5］的许可复制）。

2.4 注

1. 光学纯（对应体纯）的非天然氨基酸可能不易购得。如果是消旋混合物，应使用20μg/ml L-对应体。如果非天然氨基酸是消旋混合物且要与天然氨基酸混合的话，应加倍非天然氨基酸的用量。例如，95%4fW需要38μg/ml DL-4fW加1μg/ml L-W[5]。

2. 为了掺入非天然氨基酸，优先使用营养缺陷型菌株，但并不是对所有应用都是

必需的。为了选择展现基因编码模糊的菌株，Döring 等使用了对缬氨酸和半胱氨酸是原养型的菌株，并选择在缬氨酸编码处可能用半胱氨酸代替缬氨酸的变种；结果是一个编码缺陷的异戊氨酰转移 RNA 合成酶[19]。

3. 此处讨论的是典型表达质粒的载体。一个是使用 T7 RNA 聚合酶系统并在 *lac* 操纵子的控制之下；而另一个是为了表达目标基因则直接使用 *lac* 启动子。许多其他的表达载体已被用于把非天然氨基酸掺入感兴趣的蛋白质中，并用于完整的、部分的和定点的类似物掺入[1,5,19]。

4. 为了达到最大的表达和感兴趣的基因的掺入可能需要优化。优化的关键因素包括接种和转移到含引导物和非天然氨基酸的培养基之间的时机，以及表达延续时间的长短。例如，发现中指数期的培养液过夜表达的 GST 达到极大，而从中指数期再延续几小时的表达对 GFPuv 是最好的。更引人注目的是，用于荧光试剂的强蓝绿荧光蛋白的表达，包括荧光激活细胞分选，诱导过夜培养液再表达 6h 的蛋白质最好（J. M. Bacher 和 A. D. Ellington，未发表）。

5. 在这些例子中，用 HPLC 对水解蛋白的分析，通过检查产物中的天然的吸收光色氨酸得到简化。如果对其他氨基酸类似物完成类似的实验，氨基酸在 HPLC 分析之前可能需要衍生化。

6. 在 0.45μm 孔径的 Millipore 旋转柱中完成实验，对胰蛋白酶处理更有效（M. P. Robertson，私人通信，2003 年 5 月 12 日）。在反应的最后时间，胰蛋白酶可以用离心来纯化。固定化的胰蛋白酶吸附在膜中，洗脱液携带反应产物。

致谢　本工作得到航空和空间部宇宙生物学研究所的项目支持，项目编号 NCC2-1055；以及军队研究办公室，MURI，DAA，项目编号 19-9-1-0207。J. M. B. 得到 Harrington 学位论文奖学金的部分支持。

参考文献

1. Wang, L., Zhang, Z., Brock, A., and Schultz, P. G. (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 100, 56-61.

2. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethion-ine in *Escherichia coli*. *Eur. J. Biochem*. 230, 788-796.

3. Bacher, J. M., Hughes, R. A., Tze-Fei Wong, J., and Ellington, A. D. (2004) Evolving new genetic codes. *Trends Ecol. Evol.* 19, 69-75.

4. Bacher, J. M., Bull, J. J., and Ellington, A. D. (2003) Evolution of phage with chemically ambiguous proteomes. *BMC Evolutionary Biol.* 3, 24.

5. Bacher, J. M. and Ellington, A. D. (2001) Selection and characterization of *Escherichia coli* variants capable of growth on an otherwise toxic tryptophan analogue. *J. Bacteriol.* 183, 5414-5425.

6. Wong, J. T. (1983) Membership mutation of the genetic code: loss of fitness by tryptophan. *Proc. Natl. Acad. Sci. USA* 80, 6303-6306.

7. Bronskill, P. M. and Wong, J. T. (1988) Suppression of fluorescence of tryptophan residues in proteins by replacement with 4-fluorotryptophan. *Biochem. J.* 249, 305-308.

8. Parsons, J. F., Xiao, G., Gilliland, G. L., and Armstrong, R. N. (1998) Enzymes harboring unnatural amino acids: mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan. *Biochemistry* 37, 6286-6294.

9. Pratt, E. A. and Ho, C. (1975) Incorporation of fluorotryptophans into proteins of *Escherichia coli*. *Biochemistry*

14, 3035-3040.

10. Zhang, Q. S., Shen, L., Wang, E. D., and Wang, Y. L. (1999) Biosynthesis and characterization of 4-fluorotryptophan-labeled *Escherichia coli* arginyl-tRNA synthetase. *J. Protein Chem.* 18, 187-192.

11. Browne, D. R., Kenyon, G. L., and Hegeman, G. D. (1970) Incorporation of monoflurotryptophans into protein during the growth of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 39, 13-19.

12. Guan, K. L. and Dixon, J. E. (1991) Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* 192, 262-267.

13. Xu, Z. J., Love, M. L., Ma, L. Y., et al. (1989) Tryptophanyl-tRNA synthetase from *Bacillus subtilis*. Characterization and role of hydrophobicity in substrate recognition. *J. Biol. Chem.* 264, 4304-4311.

14. Hendrickson, T. L., Nomanbhoy, T. K., de Crécy-Lagard, V., et al. (2002) Mutational separation of two pathways for editing by a class I tRNA synthetase. *Mol. Cell.* 9, 353-362.

15. Bacher, J. M., de Crécy-Lagard, V., and Schimmel, P. (2005) Inhibited cell growth and protein functional changes from an editing-defective tRNA synthetase. *Proc. Natl. Acad. Sci. USA* 102, 1697-1701.

16. King, P. V. and Blakesly, R. W. (1986) Optimizing DNA ligations for transformations. *Focus* 8, 30-32.

17. Minks, C., Alefelder, S., Moroder, L., Huber, R., and Budisa, N. (2000) Towards new protein engineering: *in vivo* building and folding of protein shuttles for drug delivery and targeting by the selective pressure incorporation (SPI) method. *Tetrahedron* 56, 9431-9442.

18. Ausubel, F. M. (1987) Current protocols in molecular biology. Greene Pub. Associates and Wiley-Interscience, New York.

19. Döring, V., Mootz, H. D., Nangle, L. A., et al. (2001) Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* 292, 501-504.

20. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292, 498-500.

3　盘绕螺旋结构的设计和优化技巧

Jody M. Mason, Kristian M. Müller, and Katja M. Arndt

概要　虽然表观上简单，盘绕螺旋（coiled coil）模体是高度专一的，并在理解三级结构及其形成方面具有重要意义。最常观察到的盘绕螺旋形态——平行二聚态，其一般的结构类型仍有待全面的描述。尽管如此，其结构已呈现出在某些特定位置需要某些特定类型氨基酸的严格规则。本章我们基于现有盘绕螺旋结构，就应用这些规则到要设计或优化的盘绕螺旋结构来讨论这个系统。基于这些规则之上的理解和扩展，对这些模体的应用是关键的。因为，这些模体实际上在每个细胞过程中都起关键作用，它们通过屏蔽其他非天然的或不规则的蛋白质（如在肿瘤形成中跟盘绕螺旋结构域结合），作为药物投递试剂。我们讨论其结构中的 a 和 d 的"疏水"核心位置以及 e 和 g 的"静电"边缘位置，在引导寡聚化和配对化的特异性作用。同样被讨论的还有那些在维持 α 螺旋倾向性、螺旋可溶性和二聚体稳定性上与 b、c 和 f 位有协同作用的位置。

关键词　盘绕螺旋；七单元重复；体内选择；亮氨酸拉链；库设计；蛋白质设计；蛋白质工程；蛋白质片段互补测试；蛋白质稳定性；理性设计

3.1　介　　绍

盘绕螺旋是常见的结构，据估计在大多数基因组中占到编码残基的 3%～5%[1]。它由 2～5 个 α 螺旋组成，通常以左旋这种特殊方式相互缠绕，形成一个超螺旋。一般规则的 α 螺旋形成一圈螺旋需要 3.6 个残基，而左旋盘绕螺旋则只需 3.5。这意味着每两圈螺旋含有七氨基酸的重复[2,3]。最常见的盘绕螺旋类型是平行（即两个螺旋并排地从 N 端盘绕到 C 端）二聚和左旋的。在这一类型中，每个螺旋的周期都是 7。在蛋白质内的任何位置都可能有 2（在设计的盘绕螺旋中；参考文献［4］）～200 个这样的重复[5]。在这个重复中，一个螺旋的残基标记为 $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$，另一个则标记为 $(a'\text{-}b'\text{-}c'\text{-}d'\text{-}e'\text{-}f'\text{-}g')_n$（图 3.1）。在这个模型中，a 和 b 通常是位于两个螺旋的交接处的非极性的核心残基，相反，e 和 g 是部分地溶液暴露的极性"边缘"残基，通过静电相互作用，它们使得两个螺旋的相互作用有特异性。最后，剩下的 3 个残基（b、c 和 f）是典型的亲水残基并暴露于溶液中。盘绕螺旋结构连同它的七元周期性在表观上的简单性，使它得到了大量研究。引人注目的是，两个螺旋间的相互作用仍然是高度特异性的过程。正是结构周期表面的简单性与高特异性和强亲和力相结合，使这个普遍存在的蛋白质结构类型这样地引人注目。

在天然的和设计的含 2 个或 3 个螺旋的标准盘绕螺旋中，界面边缘的互补电荷可缓解不同螺旋间的排斥这就是以促进杂寡聚的形成。我们把这称为肽拉链（PV）假定[8]。这是在参考了 Kim 及其合作者设计的、他们称之为"Peptide Velcro"的、专性

杂二聚盘绕螺旋[9]之后提出的（Velcro 是尼龙的商品名，由两个法文单词“ve-lour”——钩和“crochet”——圈连成；是由钩和圈组成的带子，轻压可以将两条带子连在一起，撕拉又可以将它们分开——译者注）。除 e 和 g 位外，这两条多肽是相同的。在 e 和 g 位，一条链是 Lys，而另一条是 Glu（见注 2）。像其他类似的多肽对一样，这一对多肽在试管中形成稳定的杂二聚体。最近的一项研究通过直接将理性设计和基因选择技术的结果进行比较，检验了 PV 假定[8]。与 PV 假定相反（但与许多天然盘绕螺旋的序列性质相符），选择出的多肽对，既没有极大化相互吸引的 g/e′电荷对，也没有消除相互排斥的 g/e′电荷对（见 3.2.2.3）。许多因素可以影响 g/e′带电残基的贡献。整体静电势——包括分子间的和分子内的相互作用，起主要的作用。与核心残基的相互作用，如有利的堆积或立体的冲撞，都被用来模拟 g/e′相互作用。序列范围内的其他效应，来自于局部的螺旋柔性或来自于与 b、c 或 f 残基的相互作用，都有可能发生。对盘绕螺旋结构的检查提示，e 和 g 位在结构上有区别，这些区别可能会以不同的方式容纳不同的电荷对。

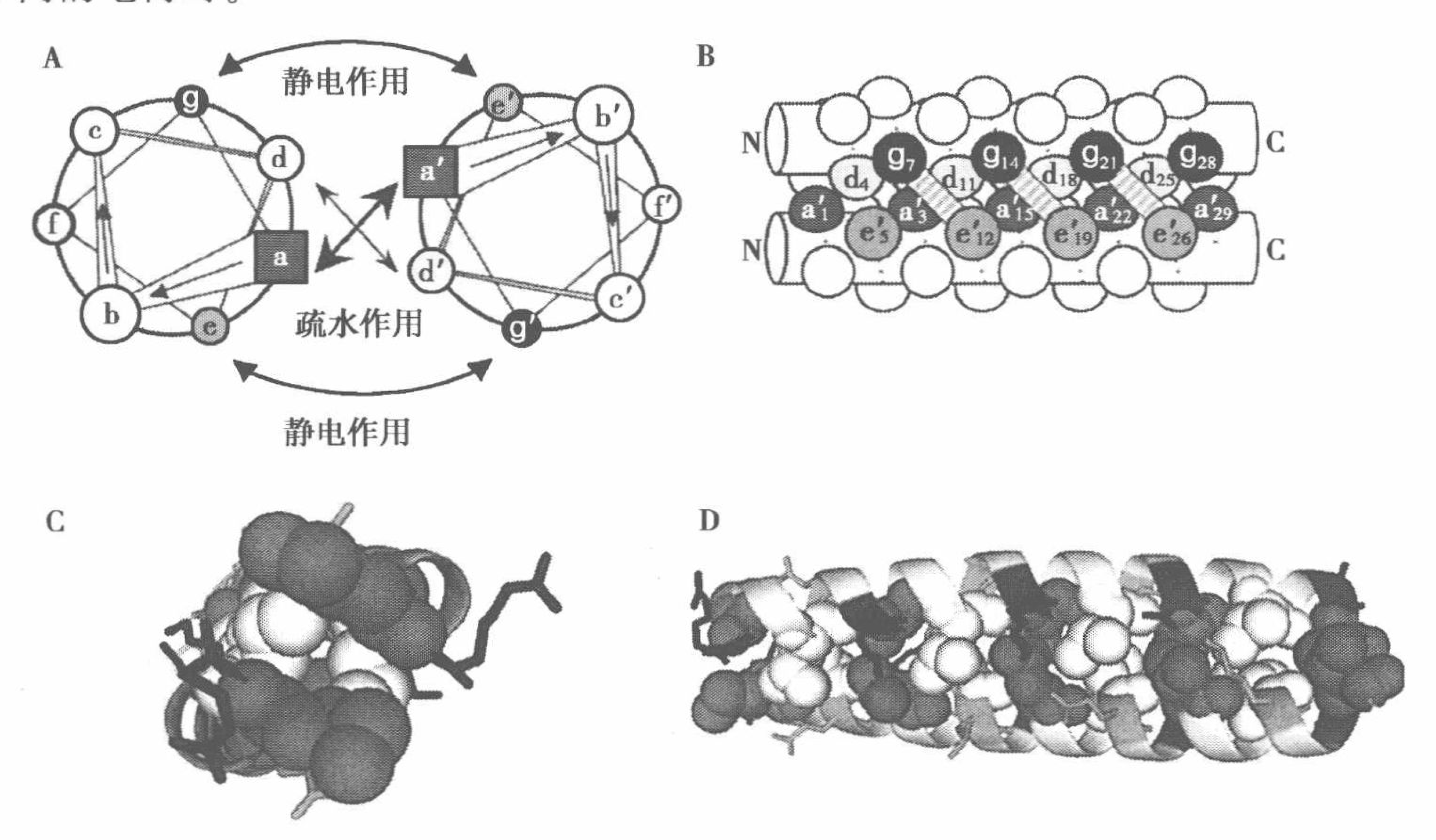


图 3.1　二聚的盘绕螺旋。A. 螺旋轮的取向是沿着螺旋轴从 N 端到 C 端。七元氨基酸位点依次分别标记为 a～g 和 a′～g′，并以不同灰度标示 a、d、e 和 g。B. 侧视图。螺旋主链用圆柱形表示，侧链用小球表示。绕在圆柱上的虚线表示肽链的通过路径。为简单起见，螺旋的盘绕没有显示出来。在位置 a（深灰）和 d（浅灰）的残基构成疏水界面，而位置在 e（中灰）和 g（黑）的残基堆积到疏水核心上。它们参与一个螺旋上的残基 i（g 位）和另一个螺旋上的残基 i′+5（e′位，属于下一个七元重复）之间的螺旋间静电相互作用，图中用短棍表示。C，D. 酵母菌转录因子 GCN4（见注 1）盘绕螺旋结构域的条带图（蛋白质结构库，PDB 索引：2ZTA；参考文献［6］）（原文索引误为 2ETA——译者注），特别显示超螺旋 g/e′相互作用。图用软件 Pymol[7]绘成。

在这里，我们对每个氨基酸在维持 α 螺旋结构合促进形成符合期望的寡聚态和左旋取向的盘绕螺旋结构中所起的重要作用做一个概述。本章的目的是强调在设计或优化这类盘绕螺旋时需要考虑的重点。这一章因而应该看作是一个“指南”，为便于盘绕螺旋的设计，解释得到期望的寡聚态（见 3.2.1）、特异性（见 3.2.2）、螺旋取向（见

3.2.3）和稳定性（见 3.2.4）所需要的最重要的操作。我们讨论在 a 和 b 疏水核心位以及 e 和 g 静电边缘位的氨基酸残基共同的影响，以及这些残基与 b 位、c 位和 f 位残基一起在保持 α 螺旋倾向、螺旋可溶性和二聚体整体稳定性上的作用。另外，也要讨论 N 端加帽和 C 端加帽的可能性。除非在正文中特殊地声明，我们都以二聚平行盘绕螺旋模体为参考态。我们还讨论了最近一篇述评中所述的盘绕螺旋在一般意义上的稳定性和特异性[10]。

对螺旋结合规律的理解使得对这些螺旋的新探索成为可能[11]。例如，通过将抗体 Fv 片段与螺旋融合产生被螺旋稳定的抗体[12]；或将抗体 scFv 片段与螺旋融合以产生微小抗体（miniantibody）[13]；或作为热感应器（如连接了绿色荧光蛋白的 TlpA，用荧光变化作为读出信号监测 TlpA 的结构随着温度变化）。这使得由盘绕螺旋二聚体形成参与的信号传导过程的测定成为可能[14]。

3.2 材料和方法

本节讨论盘绕螺旋特异性设计所涉及的几个不同方面。我们的目标是在核心处和边沿位置选择氨基酸以得到期望的寡聚态（见 3.2.1）、特异性（见 3.2.2）和螺旋取向（见 3.2.3）。这里，我们也把针对特定稳定性的不同设计方案联系起来。第 4 小节（见 3.2.4）涉及整体稳定性，集中在外部位点（b、c 和 f）。本章我们把所有内容分为小节逐一讨论。但是，正是由于所讨论的导致盘绕螺旋形成的相互影响的各因子的本性，本节的讨论会在几个地方互相交叉。因此，对于处理同一残基位置所起的不同作用的小节可以被认为是互相联系的片段。

3.2.1 寡聚态

为正常地发挥功能，任何蛋白质都必须折叠成合适的三维结构。类似地，盘绕螺旋结构必须采取正确的寡聚（四级）结构。在下一小节，讨论实现正确结构的主要因素。

3.2.1.1 核心残基

见图 3.1，我们称 a 和 d 残基为核心残基，因为它们绕着每一个螺旋形成疏水带。a 和 d 重复残基的非极性本性，有利于沿着每一螺旋的一面发生寡聚。这与在球蛋白折叠时蜷缩成的疏水核心类似，并代表对盘绕螺旋整体稳定性的主要贡献。结果是，核心残基在定义寡聚态上发挥主要的作用。

位置 a 和 d 上的疏水侧链，以“杵适配臼”（knob-into-hole）的方式掩埋于相邻螺旋中，并最先由 Crick 在 1953 年所描述[15]。在这一模型中，一个 α 螺旋的一个侧链（杵），插入对面的 α 螺旋上的 4 个侧链围成的空间（臼）中，反之亦然。堆积几何被 C_α—C_β 键与臼底部的 C_α—C_α 向量构成的角在螺旋底部的投影所定义。对平行二聚体而言，位置 a（七元重复位置 i）的杵，适配到另一个螺旋的由位置 a'_i、d'_i、g'_{i-1} 和 d'_{i-1} 上的残基顺时针排列（如果对着空洞看去）而成的臼中。因而，d_i 位的杵是在 d'_i、a'_i、e'_i 和 a'_{i+1} 围成的臼中（图 3.1）。

对 GCN4（Jun 转录因子的酵母同源物，有时称为 GCN4-p1，见注 1）盘绕螺旋不同核心突变的全面而完整的分析，揭示了不同寡聚态的不同堆积几何（表 3.1；参考文献 [16]）。GCN4-p1 二聚体 X 射线结构中的侧链堆积，和一个设计的四聚 GCN4 突变体的比较，证明 a 层和 d 层的局部几何被翻转了。平行的杵适配臼堆积在二聚体的 a 层和四聚体的 d 层。相反，垂直的杵适配臼堆积，在二聚体的 d 层和四聚体的 a 层。在平行的三聚变体中，在 a 和 d 位发现了第三类的杵适配臼堆积[17]，在这两层中，每个杵的 C_α—C_β 键与在对应臼底部的 C_α—C_β 键成近似的 60°角。这一安排称为“尖锐”的杵适配臼堆积。

表 3.1 GCN4-p1 核心突变对寡聚态的影响

位置 a～d	a～d 侧链的几何结构	寡聚状态	T_m[a]
GCN4-p1[b]	β-γ	二聚体	53℃
I-L	β-γ	二聚体	>100℃（77℃）
I-I	β-γ	三聚体	>100℃（70℃）
L-I	γ-β	四聚体	>100℃（94℃）
V-I	β-β	—[c]	73℃
L-V	γ-β	三聚体	81℃
V-L	β-γ	二聚体和三聚体混合物	95℃
L-L	γ-γ	三聚体	>100℃（76℃）

源自参考文献 [16]。

a. T_m测量在含 150mmol/L 氯化钠和 10μmol/L 肽的 50mmol/L，pH 7.0 的磷酸缓冲液中完成。括号中是在 3mol/L 盐酸胍中的测量值。

b. 野生型-p1 不同于 V-L 突变体之处在于，保证二聚化的中心 a 位被一对 Asn 替换（见 3.2.1.2）。

c. 未能确定的寡聚肽。

这些不同的几何关系，解释了一定的寡聚态有独特的氨基酸偏好。下面列出几个实验结果，这些实验测试不同氨基酸对稳定性和寡聚特异性的影响。

（1）特异的疏水残基在强制盘绕螺旋采用何种寡聚态上是关键的。Harbury 等系统地把 GCN4（除 a1 Met 外）所有的 a 和 b 残基改变为 Leu、Val 或 Ile（见注 2；参考文献 [16]）。如表 3.1 所示，这使得盘绕螺旋采用不同的寡聚态。GCN4-IL（IL 指 a 位的 Ile 和 d 位的 Leu）、II 和 LI 突变体分别是二聚、三聚和四聚，并且在研究的整个浓度范围内（寡聚态）与浓度无关。VI、VL、LV 和 LL 突变体以多种方式寡聚。L、V 和 I 的每一个组合给出独特的堆积偏好，因而形成独特的几何特征。

a. 带 β 分支的残基（Val 和 Ile）适合于平行的杵适配臼堆积（二聚体的 a 层和四聚体的 d 层），而 γ 分支的残基 Leu 适合于垂直的（二聚体的 d 层和四聚体的 a 层）几何特性。相反，β 分支的氨基酸插入垂直位，需采用热力学不利的旋转构象态[18]。

b. 很有可能，Ile 和 Val 虽有相似的立体化学，但在建立寡聚态上是不等价的。在 a 位的 Val 导致二聚体的能力比 Ile 弱。与 Val 导致二聚-三聚混合物相反，Ile 给出较高的二聚特异相互作用[17]。

c. 与二聚体和四聚体结构相反，三聚体内部的堆积，在两个疏水位点，可以适应具有最适旋转构象态的 β 分支残基。

(2) 在 Woolfson 和 Alber 的研究中，包括了对核心残基作用的研究，并把这些作用用来衡量七元重复残基的二聚和三聚倾向[19]。分析二聚体和三聚体的特性，并用这些特性来鉴别新的序列。掩埋的 Leu、Ile、Asn、Lys 和 Gln 在其预测算法中是关键的，称之为 COILER。总的来说，使用了 21 个已知形成二聚体和三聚体的不同的蛋白质，相当于数据库中的 721 个七元重复。

a. 在起先考虑的 7 种氨基酸中（Ala、Phe、Ile、Leu、Met、Val），只有 a 位的 Ile 和 Leu 以及 d 位的 Leu 在统计意义上被观察到。二聚体偏好 a 位富集 Ile 和 d 位富集 Leu，而二聚体中的 d 位几乎没有 Ile。很清楚，在这些位置上对氨基酸的这些选择使二聚体的形成有最好的堆积几何［见 3.2.1.1 中（1）和表 3.1］。

b. 在盘绕螺旋序列的核心部位，Val 比 Ile 分布得更均匀。在二聚体和四聚体的 a 位和 d 位，Val 出现的频度甚至低于随机分布的预期值。这些结果与观察到的下述事实相符：a 位和 b 位的 Val 很难区分盘绕螺旋的二聚态和三聚态［第 3.2.1.1 中（1）和表 3.1］。

c. 三聚体中 a 位和 d 位的堆积是差不多的，这些位置对氨基酸少有特异性，使得疏水氨基酸的分布更均匀[16,19]。

(3) 在研究设计好的反平行盘绕螺旋核心位置的丙氨酸的位置效应时（见注 3），Monera 等发现，当丙氨酸残基在适当位置（即在同一个环上）时，会形成二聚体[20]。如果丙氨酸残基不同步，会形成四聚体。对此，最可能的解释是，四聚体中同步丙氨酸形成的孔穴高度地不稳定，因而倾向于形成二聚体，而 Leu-Ala 重复使孔穴分布在较大范围内，从而减少在疏水掩埋和范德华相互作用上的损失。这证明了寡聚特异性是核心残基堆积引起的。

(4) 软骨寡聚基质蛋白质（COMP），属于血小板反应素家族，包含 5 链的非常稳定的平行 α 螺旋盘绕结构。46 个氨基酸长的盘绕螺旋区域（见注 4）包含环分子间（即螺旋与螺旋间）的半胱氨酸二硫键[21]。五聚界面展示出杵适配臼堆积。a 位、d 位、e 位和 g 位残基构成若干杵，它们堆积入相邻亚单位的 a′-g′位、d′-e′位、c′-d′位和 a′-b′位侧链围成的臼中。只有 f 位残基是完全暴露的，而其他 6 个位置的残基是明显被掩埋的。GCN4-pLI 突变体［见 3.2.1.1 中（1）］和五聚的 COMP 结构都含有一个大的轴向孔洞。在四聚体中的隧道大小为 1.0～1.3Å[16]，并因此而排除了水分子（半径 1.4Å）。相反，沿着五聚体的小孔有几个小分子，这与其隧道的半径较大（2～6Å）一致。

(5) 把 d 残基改变为性质上更疏水的非天然氨基酸（三氟亮氨酸和六氟亮氨酸）的实验显示稳定性增加（见注 5；参考文献［22］和［23］）。

(6) 用单甲基、双甲基和三甲基化的二氨基丙酸（dap）类似物研究了 a/d 界面上的疏水掩埋，结果显示疏水性增加[22]。甲基加入一个单体的第 16 位（在类肽的第 16 位用天冬氨酸）时，使随后得到的杂二聚 GCN4-p1（见注 6）更稳定，可能是由于折叠状态的范德华相互作用得到增加和降低了折叠的溶解损失。但是 3 个甲基的加入降低了稳定性，可能是由于增加了立体化学的位阻。奇怪的是，2 个甲基加入合成的 dap 后，引起了齐三聚（相同蛋白质或多肽的结合称为齐聚或同聚——译者注）。这证明了即使是小的尺度和疏水性的变化也能够改变稳定性与折叠倾向。

简而言之，在二聚的盘绕螺旋中，核心位置氨基酸的最好选择可能是 d 位 Leu、a 位 β 分支的 Ile（或 Val）。如果使用了 Val，Asn（如同在天然盘绕螺旋中普遍见到的）

就应该相应地放在中心 a 位，以增加相互作用的特异性（见 3.2.1.2）。三聚体的最好设计是核心位都用 Ile，而四聚体最好在 a 位用 Leu，d 位用 Ile。对这种 β/γ 侧链分支（残基）安排的偏离，会导致不利的旋转异构能，降低所期望的结构稳定性，并可能产生盘绕螺旋的多种寡聚态和反平行结合态的混合，也就是使得特异性降低。

3.2.1.2 极性核心残基

与核心残基整体上疏水相反，大约 20% 的核心残基是带电的[25]。这些带电残基通常起到使结构形成正确寡聚态的作用，大概是通过保证螺旋在结构中按需要的方式排布，否则结构将以二聚、三聚和（或）四聚，以及平行和反平行混合物的形式存在。但是，通常在特异性上的所得，总伴随着稳定性的降低。以下列出这些残基在特异性和稳定性上的作用。

（1）在关于二聚体和三聚体的统计分析中［见 3.2.1.1（2）］，Woolfson 和 Alber[19] 观察到：

a. 二聚体中 Lys 和 Asn 偏向于在 a 位，而三聚体中则基本没有。在二聚体的 a 位找到 Asn 的可能性比在三聚体中大 3 倍。

b. 三聚体的 a 位富含 Gln，Ser 和 Thr 则富集在 a 位或 d 位。

c. 常常与侧面 e 位和 g 位上的谷氨酸一起，在 a 位找到掩埋的赖氨酸。例如，这可以在 Jun/Fos 杂二聚体[26] 以及 GCN4 Asn6Lys 突变体的 X 射线结构中找到［见 3.2.1.2（5）；参考文献［27］］。

（2）掩埋的 Asn 对（在处于合适的位置时），可能通过两个螺旋的两个 Asn 侧链之间生成的氢键，给二聚体以特异性，这一现象确实在 X 射线[6] 和核磁共振[28] 研究中被观察到。不满足 Asn 侧链的氢键位能的其他构象因而在能量上是不利的。

（3）如果 GCN4-p1 核心 a 位的 Asn（见注 2）被替换为 Val，以二聚化的特异性为代价，盘绕螺旋的稳定性大量增加。因为与 Ile a 位的 Val 使得特异性缺失，Harbury 等报道得到了二聚体和三聚体的混合物［见 3.2.1.1（1）和表 3.1；参考文献［16］］；同时，Potekhin 等报道了三聚体生成[29]。

（4）在另一个例子中，平行杂聚肽（见注 1）的核心 Asn 对突变为 Leu 对（产生 Acid-pLL 和 Base-pLL 多肽），得到平行和反平行四聚体混合物[30]。

（5）在 GCN4 的情形（见注 2），Alber 组把核心 a 位的 Asn 对改变为 Gln 和 Lys，以研究这是否也能得到寡聚特异性。Lys 产生的二聚体与野生型相似，而 Gln 产生二聚体和三聚体的混合物[27]。他们推论为，极性基团制约的结构唯一性，不只是由极性掩埋引起的，也依赖于侧链与环境的正确相互作用。这些与上下文有关的效应，比起仅只是给定的七元重复位置的残基频度来，预测困难要大得多。

（6）当用一个蛋白质片段的互补试剂（见 3.2.5.3）二氢叶酸还原酶选择杂二聚的盘绕螺旋时，Arndt 等发现，在一个是 Val-Leu 的核心处，相比于 Asn-Val 或 Val-Val 组合 Asn 对构成的核心更受青睐（见注 7；参考文献［31］）。这与许多天然的盘绕螺旋相符合。

（7）在另一个研究中，二聚的 GCN4-pVL［见 3.2.1.1（1）］的 a 位或 d 位被分别突变为极性残基 Asn、Gln、Ser 或 Thr（见注 8；参考文献［25］）。只有 a 位的 Asn 残基对和 d 位的 Thr 残基对能导致正确的寡聚态。可能是掩埋这些残基在核心中所损失的

溶解能，被它们间的相互作用能所弥补，因而也能保证正确的螺旋安排。

（8）堆积环境的差别使疏水残基有不同的倾向性，即使是在 a 位和 d 位。在两个应用 a 位为 Val、d 位为 Leu 的模型多肽的透彻研究中，中心的 a 位和 d 位被系统地改变为每一种氨基酸，以测试它们在稳定性和寡聚态上的作用（见注 9；参考文献［32］和［33］）。这些改变，是有关疏水核心侧链替换对双链盘绕螺旋稳定性影响的第一个透彻的定量评估，并使得为 a 位和 d 位 19 种天然氨基酸建立热力学相对稳定性标度成为可能。表 3.2 列出了那些如果放在 a 位或 d 位会导致明确的寡聚态的氨基酸[32]。A 位的 Leu-、Tyr-、Gln-和 His-取代类似物无一例外地是三链的，而 Asn-、Lys-、Orn-、Arg-和 Trp-取代类似物无一例外地形成两链单体。在中心 d 位替换的时候，Ile 和 Val（β 分支残基）导致三链寡聚态［见 3.2.1.1（1）］，而 Tyr、Lys、Arg、Orn、Glu 和 Asp 产生双链态。

表 3.2　使用 a 位为 Val 而 d 位为 Leu 的模型多肽，把中心 a 位和 d 位系统地改变为每一个氨基酸的实验结果[a]

	位置 a		位置 d	
	寡聚态[b]	规范化的稳定性[c]	寡聚态	规范化的稳定性
Leu	三聚体	100	（69% 二聚体）	100
Ile	（61% 二聚体）	105	三聚体	89
Val	（57% 三聚体）	108	三聚体	63
Tyr	三聚体	74	二聚体	67
Trp	二聚体	55	（78% 二聚体）	47
Gln	三聚体	41	（61% 二聚体）	56
Asn	二聚体	56	（72% 三聚体）	41
Lys	二聚体	37	二聚体	25
Orn	二聚体	10	二聚体	7
Arg	二聚体	31	二聚体	9
His	三聚体	28	（55% 二聚体）	37
Glu	（54% 二聚体）	10	二聚体	12
Asp	二聚体[d]	—	二聚体	24

a. 显示的是导致明确寡聚态的氨基酸（采自参考文献［32］和［33］）。

b. 在检查寡聚态时，螺旋是通过 N 端 Cys-Gly-Gly 连接子用二硫键桥接的（见注 9）。

c. 规范化的稳定性是设定稳定性 Gly＝0 和 Leu＝100 后每个取代类似物的相对稳定性。

d. Asp 类似物在 5°C 100%的折叠。在室温，此类似物只有约 20%折叠。

（9）Ji 等突变了 gp41——来自猿猴免疫缺陷病毒的 6 螺旋束包膜蛋白，与 gp120 一起，负责病毒与 $CD4^+$ 细胞的融合[34]。在结构上，它是由反平行杂二聚体组成的三聚体蛋白。在这一研究中，为核心氢键和盐桥负责的（两个 Gln 和两个 Thr 残基）4 个被掩埋的极性残基的每一个，都单独地突变为 Ile。其中，3 个形成更稳定的 6 螺旋束，另 1 个形成不溶的聚集体（见注 10）。这些结果证明了这些残基具有控制结构稳定性和特异性间平衡的能力。这是非常重要的，因为在融合之前，此蛋白质要经受结构变化，并必须在两个结构之间有正确的稳定平衡，以使得这种变化是可能的。这些极性核心残基协助调控此构象稳定性，因而，也协助膜融合本身。

一般来说，核心 a 位的 Asn 残基对显然在二聚体中占支配地位，特别是，如果核心偏离最优的 Ile-Leu a-d 残基布置的时候，因为这一组合似乎导致平行二聚，而不需要核心内的 Asn 对。核心的 a 位 Gln 对三聚体可能是好的选择，尽管为了确保三聚体排它的特异性，可能还需要其他一些因素。

3.2.1.3 边沿残基

七元重复的 e 位和 g 位（边）在盘绕螺旋界面上从侧面围着 a 残基和 d 残基（图 3.1）。这些位置的掩埋，高度地取决于寡聚态。结果，在 e 位和 g 位氨基酸的选取可能受到寡聚态的影响。表 3.3 显示掩埋面积的百分数，表示在螺旋单独存在时的侧链可及表面积在寡聚态中变成掩埋状的分数[16]。

表 3.3 GCN4-p1 二聚体和 p-LI 四聚体中掩埋表面积的百分数

位置	GCN4-p 1 二聚体	GCN4-pLI 四聚体
a	87	92
b	0	10
c	1	19
d	87	99
e	26	72
f	0	0
g	27	66

采自参考文献 [14]。

（1）与二聚体中对应位置相比较，三聚体的 e 位和 g 位富含疏水残基（Ile、Leu、Val、Phe、Tyr 和 Trp）并几乎没有特异的亲水残基（Glu、Gln、Ser 和 Arg；参考文献 [19]）。这些模式是与三聚体中疏水界面相对于二聚体的扩张相符的。疏水残基百分数的增加使狭窄的疏水面的宽度增加，并且高寡聚态的似然率也增加了。在高寡聚态中有比两螺旋的盘绕螺旋更多的非极性掩埋。这可以从表 3.3 中看出。在 e 位和 g 位的疏水掩埋百分比，大约增加了 40%。三聚体中在对面带电的 g_i 到 e'_{i+1} 残基对（12%）相对于二聚体（23%）的减少与此相符[19]。

（2）Fairman 等突变了 lac 阻遏蛋白的 C 端同四聚盘绕螺旋结构域，以产生杂四聚盘绕螺旋[35]。在 e 位和 g 位周边的 b 位和 c 位，全为 Lys 或全为 Glu 的多肽，结合力很弱。但是，把它们混合起来，就会形成高度稳定的四聚体（见注 11)。这证明了，至少对于四聚体，b 和 c 残基在盘绕螺旋的稳定性上也起了重要作用。这种作用类似于二聚盘绕螺旋中存在的 g/e′离子相互作用，但四聚体加宽了的疏水界面扩展到这些残基，离子作用从核心开始向外后退到 b 和 c 残基。通过改变 pH 和盐的水平，Glu 和 Lys 间的这些离子对相互作用，被证明为对稳定性的增加负责。另外，这些电荷在齐性寡聚时直接相对，并且很可能在潜在的齐二聚相互作用体中，这一不利的电荷排斥推动了杂二聚体的形成[9,35]。

3.2.2 配对特异性

本节讨论配对特异性的重要性，以及盘绕螺旋可以用来保证没有其他能量上有利

的结构可以被接受的某些方法。引人注目的是，与序列和结构上的相似性相反，盘绕螺旋偏向于与功能上的伙伴相互作用。本节分析介导这种高特异性的因素。

3.2.2.1　核心残基

如 3.2.1.1 小节所述，疏水残基（主要是 Leu、Ile 和 Val）分布模式是螺旋间结合的主要推动力。但是，因为这种模式非常常见，螺旋怎样能同时用其核心抗拒不同结构的形成？答案是一个复杂的图像，包含微妙的变化，如非稳定化的、非疏水的核心残基的插入，这种变化会选择对抗另外种类的结构。

(1) Sharma 等设计了一个多肽（anti-APCp1），其目的是用于结合来自腺瘤性结肠息肉（APC）的肠癌有关肿瘤抑制蛋白的盘绕螺旋序列（见注 12；参考文献 [36]）。这样，他们采用了核心改变以及结合 g/e′相互作用，而不是 Asn 配对，以证实特异性可归因于这个相互作用。他们推理认为，考虑到核心残基在结合过程中的主导作用，以及核心突变对稳定性和特异性都有大的影响，发现在主导特异性上对核心残基的低要求是令人奇怪的。为了处理这个问题，他们设计了 1 个多肽，以结合到 APC 的头 55 个氨基酸（APC55）上，并基于角蛋白的类型Ⅰ和类型Ⅱ杂二聚体在位置 a 和 d 的共变模式，突变了此 anti-APCp1，以产生更常见的 a-a′和 d-d′配对。他们做了 3 个突变（在 a 层的 A41Ⅰ、d 层的 A2M 和 M44A），以把野生型的 Ala-Ala 和 Met-Met 相互作用分别变为更常见的 Ala-Ile、Ala-Met 和 Met-Ala 相互作用。两个更进一步的突变（a 层的 T6G 和 d 层的 N30H）用对应的 Gly-Gly 和 His-His 对来降低齐二聚体的稳定性。在定向抗 anti-APCp1 齐二聚化的同时，额外的 e-g′配对优化了离子相互作用，所得二聚体具有稳定性及特异性。

(2) Schnarr 和 Kennan 通过核心疏水残基的立体化学配合形成了杂三聚蛋白[37]。在他们的研究中，用各种不同侧链长度的非天然氨基酸来启动特异的杂三聚体形成。研究者们把 GCN4 核心 a 位残基用 Ala 或环己基丙氨酸替代（见注 13），结果得到的是三聚体中立体化学不匹配的核心层，以 3 个 Ala 或 3 个环己基丙氨酸，产生立体规避或排斥。但是，2 个 Ala 和 1 个环己基丙氨酸产生了 1 个立体化学配合很好的杂三聚体。非天然侧链可以以这种方式产生盘绕螺旋。环己基丙氨酸的额外体积弥补了 Ala 核心层，以提供立体配合，而更大的侧链只会使分子不稳定。

3.2.2.2　极性核心残基

在 3.2.1.2 讨论了极性核心残基在引导盘绕螺旋特异的寡聚态上的作用。在 3.2.1.1 曾提到，在盘绕螺旋配对中的异型核心接触允许产生杂特异性。更多的特异性可通过极性核心与外沿残基的相互作用得到。

(1) 在天然二聚盘绕螺旋中，a 位的 Lys 是仅次于 Asn 的最普通的掩埋极性残基[19]。A 位 Lys 可以与前 1 个七元重复的 e 位残基[27]形成螺旋内静电相互作用，以及与平行二聚体中，对面螺旋前 1 个七元重复的 g′位的极性残基，形成螺旋间 g′-a 极性相互作用[26]。

(2) Campbell 和 Lumb 在肽拉链（PV）的 Base-pLL 中放入了两个 a 位 Lys，以在肽拉链的 Acid-pLL 中实现螺旋间的由这 2 个 Lys 与 e′和 g′位的 Glu 残基提供的极性相互作用（见注 14；参考文献 [38]）。正如所期待的，结果偏向于二聚态，很可能是因

为在高寡聚态中会有较大的溶解损失。另外，这样的相互作用，相比于 Asn-Asn（a-a′）相互作用，对稳定性的损害较小，主要是因为后者有较大的溶解损失需要付出。然而，平行和反平行的皆有可能，主要是因为，平行取向中的 a-g′相互作用和反平行取向中的 a-e′相互作用在能量上是相似的。

（3）Harbury 的组采用计算方法［见 3.2.5.4（4）］，不仅考虑期望的结构，而且考虑不同的不期望的结构，以实现特异性[39]。改变了 GCN4 的中心七元重复的 a 位、d 位、e 位和 g 位残基（见注 15）；所有非脯氨酸残基以及同型和异型序列都经受了计算（*in silico*）选择和实验验证。次于 g/e′残基对的体积互补和电荷互补，发现 d 位的 Glu 宁愿与 e′位的 Arg 配对。

3.2.2.3 边缘残基

配对特异性受到 e 位和 g 位残基静电本质的影响（在平行二聚螺旋中，一个七元重复的 g 位与另一个螺旋的下一个七元重复的 e′位，被记为 $i \rightarrow i'+5$）。这些残基常常分别是 Glu 和 Lys。这样互补的残基对相互作用，分别增加并巩固原来由核心疏水作用提供的特异性和稳定性。盘绕螺旋外接触边缘的电荷模式会指导蛋白质是形成齐聚还是杂聚蛋白，以及盘绕螺旋内的取向是平行的还是反平行的。但是，PV 假定（见第 1 节；参考文献［8］）过分地简单化；那些在复合物稳定性上很少有或没有作用的残基，通过指导螺旋避开，那些会危及分子的特异性的，同样的或不期望的相互作用来达到其目的（负设计）。另一方面，某些盘绕螺旋似乎没有经受到进化压力，因为它们已经是特异性的，并且也不需要比它们现在的状态更稳定。

（1）如所期待的，用排斥的 Glu-Glu 配对取代有利的 g/e′位 Gln-Gln 配对已被证明使盘绕螺旋构象不稳定[40]。

（2）Hodges 及其同事估计 g/e′配对盐桥对盘绕螺旋稳定性的贡献是 0.37kcal/mol（见注 16；参考文献［41］）。

（3）细心安排 e 和 g 位电荷，可容许杂二聚体的生成，并另外保证，如 PV 假定所包含的，不倾向于生成齐二聚体[9,42,43]。

（4）在 e 位和 g 位找到的 4 个最常见的氨基酸是 Glu、Gln、Arg 和 Lys[44]。这些残基含有长疏水侧链，能够与 a 和 b 核心残基作用，并终止于一带电（Glu、Arg 或 Lys）或极性（Gln）基团。通过先突变 e，而后 g，最后两个七元重复对上的残基为 Ala，Krylov 等针对鸡的卵黄蛋白原结合蛋白成功地建立了这些相互接触残基的耦合能（$\Delta\Delta\Delta G_{int}$）[44,45]（见注 17）。这一双突变热力学循环被用来为外沿残基接触的倾向建立热力学标度（表 3.4）。在 150mmol 氯化钾和 pH 7.4 的条件下，Glu-Arg 吸引被证明比 Glu-Lys 吸引稍微稳定，合理的解释是，Arg 的侧链较长并较好地与谷氨酸相互作用，而且，各个甲基更有效地把核心与溶剂屏蔽开。这样的后果是增加带电末端基团的作用，在较弱的水性环境中产生较大的贡献。如所预期的，高盐像低 pH 一样，弱化这些相互作用。在高盐或低 pH 条件下，极性作用减弱，疏水作用增强。Glu-Arg、Glu-Lys 和 Gln-Gln 相互作用是最具稳定能力的（与取向无关）。而 Glu-Glu、Arg-Arg、Arg-Lys、Lys-Arg 和 Lys-Lys 这些相同电荷的相互作用相当地不受欢迎。

表 3.4　用双突变丙氨酸热力学循环法计算的 g-e′配对耦合能（$\Delta\Delta\Delta G_{int}$）

g-e	Glu	Glu	Arg	Lys
Glu	+0.7±0.2	+0.2±0.1	−0.5±0.1	−0.3±0.15
Glu	+0.2±0.1	0.0±0.1	+0.3±0.2	+0.3±0.1
Arg	−1.1±0.1	+0.4±0.1	+0.8±0.1	+0.8±0.1
Lys	−0.9±0.1	+0.3±0.1	+0.6±0.1	+0.6±0.1

采自参考文献［44］。单位：kcal/mol。

（5）Arndt 等设计了一个多肽库。此库的设计基于 Jun-Fos 杂二聚，库中 b、c 和 f 残基来自于各自的野生型蛋白，a 位和 d 位为 Val 和 Leu（带有 a3Asn 在核心的插入例外，此插入引导期望的螺旋取向和寡聚态），e 和 g 残基则用三核苷酸作改变以得到等摩尔的 Arg、Lys、Gln 和 Glu 混合物（见注 7；参考文献［8］）。出人预料的是，最好地选择出来的成功者——Winzip-A2B1 杂二聚体（见注 18），在 g/e′位缺少完全互补的带电残基，尽管选择过程是很彻底的。更确切些说，6 个 g/e′对的两个预测为排斥的，提示得到的序列偏离简单的电荷互补规则（PV 假设）。可以假定，整体静电势（包括分子内和分子间相互作用）起主要的作用，并且与核心残基的作用，如像有利的堆积或立体冲撞，也可以调节这些 g/e′作用（见参考文献［8］和［31］及其中引用的文献）。这样的观察与天然发生的盘绕螺旋相符，这些天然结构通常有复杂的相互作用模式。这些盘绕螺旋必须满足一些标准，如生物稳定性和在家族内的极高特异性，而与其他盘绕螺旋家族间没有交叉反应[46]。

3.2.3　螺旋取向

盘绕螺旋多数折叠为平行排布，但是，越来越多的测定了结构的蛋白质含有反平行的盘绕螺旋结构域[47]。与反平行盘绕螺旋的生物学重要性逐渐得到承认相反，对这类分子的研究被没有行为规范的模型体系所牵制。例如，在像丝氨酰转移 RNA（tRNA）合成酶或丁型肝炎抗原那样的蛋白质中发现的反平行盘绕螺旋，都被证明足以实现没有更进一步结合的二聚化。

Hodges 及其合作者第一个报道了全新设计的，被内部二硫键限定在反平行取向的盘绕螺旋的表征[48]。这一以及其他设计的反平行盘绕螺旋，分别比它们的有几乎等价的螺旋间相互作用的平行对应物更稳定[49,50]。这些数据显示，假定其他一切都是等价的，螺旋偶极相互作用（见 3.2.4.3）偏好反平行取向。

3.2.3.1　核心残基

在平行盘绕螺旋中，互相堆积的核心残基有 a-a′和 d-d′。在反平行盘绕螺旋中有 a-d′和 d-a′中心堆积，产生同样的堆积层[51]。

（1）已经证明，Ala 残基在全新设计的盘绕螺旋核心的相对位置可以控制平行或反平行取向［见 3.2.1.1（3）；参考文献［52］］。通过细心地放置 Ala 在中间的七元重复中，全平行的或全反平行的四聚体都可以形成。在分子中间核心位置的七元重复的中间两层平面用另一种不同的残基对 Ala 和 Leu（Ala-Leu-Ala-Leu），也可以达到同样目的。这样由大小两种残基交替组成的核心（Ala-Leu-Ala-Leu）是引入小侧链的最好办法。在平行安排中，堆积会是在一个平面中都是 Ala，而另一个平面中都是 Leu，于是导致一个大空洞从而溶解核心并使分子不稳定（见注 19）。

(2) 在一个类似实验中，二聚的 GCN4-p1（见注 2）核心 Asn 被变为 Ala。结果是得到反平行三聚体，以避免核心空洞[53]。但是，在苯存在的条件下，得到的是平行三聚体，苯结合在核心空洞中[54]。

(3) 在最近称为 APH 的反平行同二聚盘绕螺旋的设计中，疏水核心中的 β 分支的立体匹配（Ile 于 d 位）和截断的侧链（Ala 于对面的 a′位）以及其他特性［也见 3.2.3.3 (3)］一起被用来启动反平行取向[55]。平行排布在 Ile 的 d 层是能量上没有优势的，从而很差地被放在二聚盘绕螺旋中（见 3.2.1.1），和疏水核心的 Ala-Ala 空洞从能量上排除（见注 20）。

3.2.3.2　极性核心残基

与平行盘绕螺旋相似，与对应掩埋的极性残基也在规定螺旋取向上起到关键的作用。

(1) 在平行盘绕螺旋中，普遍地观察到 a-a′的 Asn 配对（见 3.2.1.2）。但是，在反平行盘绕螺旋中，对应的 a-d′位并未观察到这种配对（综述见参考文献［47］）。确实，在反平行盘绕螺旋中的 Leu-Leu 相互作用比在平行的核心堆积中更稳定[49]，并且很可能是 Asn-Asn 相互作用和静电相互作用（见 3.2.2.3 和 3.2.3.3）导致平行构象相对于反平行构象的特异性。

(2) 尽管在天然蛋白中没有 Asn-Asn a-d′配对，Oakley 和 Kim 修饰了平行杂二聚盘绕螺旋肽拉链，使得仅当螺旋在反平行取向时，才可以存在掩埋的极性相互作用，(即 a-d′配对；参考文献［56］)。据估计，这一掩埋的极性相互作用给予约 2.3kcal/mol 中等程度的反平行偏向（见注 21）。但是，仅在掩埋的极性相互作用与引进平行取向时电荷排斥两者结合时，才能排它性地形成反平行盘绕螺旋［见 3.2.3.3 (2)］。

(3) 与在平行二聚盘绕螺旋中存在的 a-g′或 d-e′相互作用可以比较（见 3.2.2.2），a-e′或 d-g′相互作用可以出现在反平行盘绕螺旋中。例如，在丝氨基 tRNA 合成酶盘绕螺旋中，一条链上 d 位的 Arg-54 和另一条链上 g′位的 Glu-74[57]，或者在 GreA 盘绕螺旋中的 a-e′相互作用[58]。据假定，这种类型掩埋的极性相互作用也可能在取向特异性中起作用。Oakley 的小组取代了肽拉链的核心 Asn 残基对，以使得在反平行取向中，Acid-p1 的 a 位 Arg 可以与 Base-p1 的 g′位 Glu 作用（见注 22；参考文献［59］)。平行构象的稳定性会被潜在的 e′位 Lys 的排斥作用降低。但是，与 Campbell 和 Lumb 的研究一致［见 3.2.2.2 (2) 和注 14；参考文献［38］］，虽然核心引进 Arg 后能够促进二聚化，但并未发现对反平行或平行取向有明显的偏好。两种状态的自由能差估计仅有 (0.1±0.1) kcal/mol。一个可能的解释是归因于与邻近的 g 位 Glu 形成了螺旋间相互作用。

3.2.3.3　边缘残基

在平行盘绕螺旋中，极性的 e 位和 g′位产生有利的作用。在反平行的盘绕螺旋中，e 与 e′作用，g 与 g′作用，结果是一个螺旋被等效地旋转了 180°。这种倾向性的改变，在取向选择上有着重要作用。

(1) Monera 等设计了平行的和反平行的盘绕螺旋，预计会有链间吸引或排斥（见注 23；参考文献［49］)。确实发现，在主要的取向中有带相反电荷氨基酸间的静电相互作用。

(2) Oakley 的小组改进了早先设计的反平行杂二聚盘绕螺旋［见 3.2.3.2 (2)］。在他们的第一个设计中[56]，在 e 位和 g 位，一条肽链仅含 Glu，另一条肽链仅含 Lys (见注 21)。在他们的新设计中，替换了每条肽链的 g 位残基，以使得所有潜在的吸引作用都预计只在反平行取向存在（见注 24；参考文献［60］)。相反，两个潜在的库仑排斥作用预期存在于平行取向中，并且确实证实了反平行排布的强烈倾向。

(3) 在最近设计的反平行齐二聚盘绕螺旋 APH［也见 3.2.3.1 (3)］，N 端 e 位和 g 位的 Glu 与 C 端 e 位和 g 位的 Lys 被用来引导反平行取向，结果在反平行排布得到 8 个潜在的库仑作用，在平行排布得到 8 个潜在的排斥作用（见注 20；参考文献［55］)。

3.2.4 稳定性

达成有利的稳定性是大量效果相反的焓和熵的净结果。结果是得到一个具中等程度稳定性的蛋白质，进化后可以在生理条件下与其伴侣蛋白相互作用，但不是稳定到不会解离的程度。达到这个平衡，是核心和边缘残基以及它们的相互作用、螺旋长度、成螺旋的倾向、溶解度和螺旋戴帽的结果。本节讨论这些稳定化因素。

一个重要而微妙的因素是溶解度。参与的螺旋必须有一非极性内核，以容许有利的相互作用，但螺旋的整体非极性必须不至于强到在工作条件下发生聚集。残基 a 和 d 必须形成一个连接两个螺旋的疏水带，e 和 g 典型地应该是极性残基，参与保证只有真正的结合伙伴与螺旋作用（即与非作用伙伴结合时变得不稳定），并增强核心导致的稳定性。这把剩下的残基留在了七元重复中的位置 b、c 和 f 上，以使电荷平衡和保证螺旋稳定且可溶。Glu 和 Lys，也有非常适合的螺旋倾向。带电残基也可与螺旋偶极有利地作用（见 3.2.4.3)，还可以与螺旋内一圈以外的带电残基有利地作用。这给予在离开二聚界面的地方，暴露于溶液的位点选择这些残基的额外优点。在溶液暴露位置插入 Tyr 有利于肽浓度测定。粗略地说，对原型的二聚盘绕螺旋，如 GCN4 和 Jun 蛋白，在这些位置 Lys、Asp、Arg、Glu 和 Asn 有优势。虽不常见，但也存在的是 Ala，也许在附近空间的残基引起整体 α 螺旋倾向降低的情况下，可以认为它给螺旋以额外的稳定性。

3.2.4.1 螺旋长度

一般来说，在盘绕螺旋链长度增加时，观察到稳定性的（线性）增加[61]。这是因为盘绕螺旋的序列将会起到额外的重要作用。例如，Lau 和 Hodges 构建了一个比原肌球蛋白（284 残基盘绕螺旋）还稳定的 29 聚体（见注 25；参考文献［62］)。在增加了的疏水掩埋、氢键以及极性相互作用的基础上，可以期待，当长度增加时，盘绕螺旋结构变得更加稳定。在对设计的齐二聚肽的长度研究中（见注 26；参考文献［63］)，这已被证明。虽然最短两个七元重复（见注 27；参考文献［64］）也是需要的，以在同三聚体的情况下容许结合。Lac 阻遏蛋白的盘绕螺旋结构域被用来作为检测链长对稳定和折叠的效应的基础[65]。毫不奇怪，四聚体的解离常数随每个螺旋中七元重复数的增加而降低（见注 28)。有些更令人惊奇的事实是，每个螺旋少到 4 个的七元重复协同地折叠，并没有二聚中间物的迹象。

与短盘绕螺旋相反，长的盘绕螺旋，如像原肌球蛋白和肌球蛋白重链结构域，并不在它们的核心排它性地富含大块的非极性氨基酸。更恰当地说，由于蛋白质长度所提供的稳定性，这样的蛋白质含有小的非极性或带电残基[66]。这些残基约占到核心的

40%。这些不利于稳定的残基集团主要由 Ala 组成，因为这种残基比起极性残基来说对核心的稳定性损害较小（比极性侧链更容易堆积），并能增加整体的螺旋倾向性。这意味着，尽管获得正比于七元重复数的稳定性增加，但并未在这些长盘绕螺旋中观察到，从进化的观点来看，这样高的稳定性增加是不必要的。作为代替，疏水稳定的残基集团为盘绕螺旋提供必要的稳定性，而使稳定性降低的残基集团却不提供稳定性，但它们确实维持螺旋结构。这些主要含 Ala 的残基集团，在这些区域增加游变性和局部的去折叠，却不影响盘绕螺旋的整体稳定性，很可能允许蛋白质行使其独特的生物功能。

3.2.4.2 螺旋倾向

据 Litowski 和 Hodges 报道，通过把 Ser 改变为 Ala（具最高螺旋倾向性的氨基酸）来增加非核心残基的 α 螺旋倾向，可以稳定整个的盘绕螺旋[67]。在它们的模型——Glu/Lys 盘绕螺旋（见注 29）中，这导致每个替换 0.4～0.5kcal/mol 的稳定化。这与早先 O'Neil 和 DeGrado 的结果相符[68]。这个数值小于对单个 α 螺旋的值，可能是因为维持盘绕螺旋涉及额外的稳定化相互作用。

一般来说，外部残基 b、c 和 f 应该与溶剂形成氢键，并且在同时，补偿潜在的不再能与之形成氢键的伙伴[69]。这些残基也负责帮助维持 α 螺旋倾向，其中每个残基都有一个构型偏好，或者稳定化或失稳化螺旋[68]。Ala 的 α 螺旋倾向已知是所有氨基酸中最高的，并且虽然所有普通的核心疏水残基（Asn 为例外）都有高的 α 螺旋倾向，溶液暴露的 Arg 和 Lys 也起一定的作用。这是因为 Lys 和 Arg 两者都是很好的氢键伙伴，并且非常适合用来改善分子的溶解度。确实，我们发现，螺旋倾向性在盘绕螺旋设计中是一个重要的因素[70]（http：//www.molbiotech.uni-freiburgode/bCIPA）。

3.2.4.3 与螺旋宏偶极的相互作用

蛋白质结构中 α 螺旋组成的统计分析揭示不同的氨基酸偏好螺旋的不同部位。特别地，带负电能力的残基（Asp 或 Glu）强烈地偏好接近螺旋的 N 端，而带正电能力的残基（His、Arg 或 Lys）稍弱地偏好 C 端[71,72]。解释极性侧链对螺旋末端偏好的模型可以分为两类。第一，因为 α 螺旋的头 4 个主链 NH 基团和最后 4 个 CO 基团不能与其他的主链基团形成 $i \to i+4$ 主链氢键，螺旋末端的极性侧链则可以代为氢键伙伴。这称为螺旋戴帽并在 3.2.4.4 节有描述。第二，带电侧链和由单个肽骨架偶极排列形成的 α 螺旋净偶极矩之间的静电“电荷-螺旋偶极”相互作用也可以使蛋白质稳定或失稳。

（1）Hodges 研究组研究了带负电的 Glu 侧链，对一个设计的、没有螺旋内和螺旋间相互作用的、齐二聚盘绕螺旋稳定性的影响的位置效应（见注 30；参考文献 [73]）。在盘绕螺旋的每一个链的近 N 端，一个取代 Gln 的 Glu 在 pH 7.0 使盘绕螺旋稳定，与电荷-螺旋偶极相互作用模型相符。相反，在螺旋中部用 Glu 替换会降低螺旋稳定性，因为在 pH 7.0 相比于 Gln，Glu 的螺旋倾向和疏水性更低。在 C 端的 Glu 被替换则更加降低盘绕螺旋的稳定性，这是因为本身带有的降稳定性能力，和与螺旋偶极矩的负极间不利的电荷-螺旋偶极相互作用的综合效果。

（2）在从两个在 4 个 a 位和 4 个 g 位含 Glu、Gln、Arg 和 Lys 残基等摩尔混合物的设计的蛋白质库中挑选杂二聚盘绕螺旋时（见注 7），Amdt 等观察到，在选出的盘绕螺旋序列中，带负电的 Glu 和中性的 Gln 在 N 端部分的富集，以及带正电的 Lys 和

Arg 在 C 端部分的富集[31]。这种偏好与假设的电荷-螺旋偶极相互作用符合得很好。

3.2.4.4　螺旋戴帽

螺旋可以标记为 N′-Ncap-N1-N2-N3-N4-mid-C4-C3-C2-C1-Ccap-C′。在这些位置中，N′、Ncap、Ccap 和 C′没有螺旋的 Ψ 和 Φ 角，并且 N1-N2-N3…C3-C2-C1 参与作为 α 螺旋特征的 $i \rightarrow i+4$ 氢键。残基 N1、N2、N3、C1、C2 和 C3 是唯一的，因为它们参与 $i \rightarrow i+4$ 骨架-骨架氢键的酰胺基只使用 CO 基（在 N 端）或者 NH 基（在 C 端）（见 3.2.4.3）。这些残基对形成氢键在螺旋结构和稳定性上有强有力的影响[74]。从 N4 或 C4 往上，所有残基都可以满足 NH 和 OH 骨架氢键条件。

(1) 在螺旋设计中，在 N 端对稳定性最具选择性的位置是 Ncap 位。对这个位置 6 个最适合的残基是 Ser、Asp、Thr、Asn、Gly 和 Pro，另外有 11 个残基（Val、Ile、Phe、Ala、Lys、Leu、Tyr、Arg、Glu、Met 和 Gln）被完全排除[75]。对 Ncap 的 6 个最适残基中，Ser、Asp 和 Thr 是最好的。

(2) Ncap 模体的一个好例子是 Ser-Xaa-Xaa-Glu（Ncap-N1-N2-N3），其中反向的侧链/主链相互作用样式（Ncap 位 Ser 的 OH 与 Glu 的 NH，以及 Glu 的羰基与 Ser 的 NH）使螺旋稳定。进一步的稳定可通过 Ser-Xaa-Xaa-Glu 模体前后的疏水残基达成[76]。Lu 等在 GCN4 的序列中引进了这一戴帽样式，因此而使盘绕螺旋达到 1.2kcal/mol 的稳定水平（见注 31；参考文献 [77]）。

(3) N1 位和 C′位很喜欢 Pro（在立体化学上是相容的，因为前面的残基具有非螺旋的骨架两面角），这确实是一个共同的螺旋终极模体，但必须避免出现在螺旋主体中，以及 C3、C2 和 Ccap 位。作为所有氨基酸中溶解度最高的 Pro 与螺旋末端的溶液暴露位是相容的[78]，并且不需要氢键受体，因为它没有骨架 NH 基团[76]。

(4) C 端戴帽模体包含骨架-骨架氢键，而不是在 Ncap 和 N3 之间观察到的侧链到骨架氢键。在 C 端骨架，氢键由螺旋后骨架基团来实现（如 Schellman 模体中的 C′和随后的 C″；参考文献 [76]）。这意味着，C 端只在 C′位需选择可以取正 Φ 角的残基，如 Gly。

3.2.5　理性设计与选择技巧

蛋白质设计的复杂本性意味着设计或预选具有高稳定性的序列是一件令人畏缩的任务。如果程序允许，宁愿生成一个相互作用库——可检查其序列是否有成功的相互作用的不同分子的集合。这可认为是半理性的设计，凭此可在库中引进特定位点的平衡变化（经常维持野生型残基不变），得到的大量分子组合可供筛选。正常情况下，在类似的设计中，要改变所有七元位为所有可能的氨基酸，将会产生大而不现实的库容（如果含 4 个七元重复，每个位点有两种可供选择的氨基酸，将会产生 $2^{28}=2.68\times10^{8}$ 库成员）。但如果只需产生一条待合成并测定结合强度的序列，则会是耗时且枯燥的，而且结果常会是令人沮丧的。采用半理性设计，可得到合理的折中，一个分子可以包含在库中，测试它是否确实是一个用于结合的优秀候选者，同时，提供多功能以产生新的、非直觉的、并常常是全新的结合伙伴。在基于细胞的系统中的选择，具有更多额外的优点，可以伴随地筛选出对该有机体内的蛋白酶敏感的序列。

3.2.5.1　简并密码子

使用简并密码子，可在希望改变的位点上编码若干氨基酸的混合密码。同样，在任

细选择要随机化的对应位点引入简并密码子，不仅可以引入期望的碱基，而且可以引入期望的氨基酸。如已经讨论过的，盘绕螺旋在不同位置对氨基酸类型有偏好。例如，e 和 g 残基常是极性且互补的（表 3.4），因而，Glu-Lys 相互作用可以替换 Glu-Arg 相互作用，对分子结构只有极少扰动（因为 Lys 和 Arg 在体积与电荷上相当相似）。在这种情况下，Arg 可以引进，而 Lys 可以保留在库中，以观察从分子整体上看，哪个氨基酸更适合。同样地，疏水核心在 a 位偏好 β-分支氨基酸，这意味着，虽然在野生型中可以期待找到 Val，Ile 也许更合适。但是，必须注意到，执着于这些偏好则是过分简单的，表面上不利的残基对设计非特异结构和甚至整体稳定结构是必需的（见参考文献[8]）。但是，一般来说，保持整体的“两元样式”将能够保证得到所有的库中成员，有合理的机会形成盘绕螺旋结构，而不是能量上有利的其他结构。盘绕螺旋库成员将像野生型结构所做的那样，掩埋它们的疏水残基，并用极性的 e 和 g 残基将这些疏水残基包围起来。对一个盘绕螺旋来说，疏水核心的维持，称为“3-4 重复”（a 到 d 和 d 到 a 分别是 3 和 4 个残基间隔）是基本的。最后，可能合成特异的三核苷酸聚合物，使得到的基因序列编码，并只编码期望的最终氨基酸序列[79]。这也有助于在容许更多期望的待筛选的变化的同时，保持低库容。对使用正规的简并寡核苷酸的简并密码子，情况不是这样的。依赖于密码子用法表，可能意味着为了有两个期望的氨基酸，另一个不期望的氨基酸必须包含在那个位置（见 3.2.5.2 和图 3.2）。也有可能，在密码子表相反两端的两个氨基酸是期望的，但是在现实中，只能用这样预先合成的三核苷酸密码子才能包括进来。这样的三核苷酸建筑模块近来已商品化（Glen Reserch，Sterling，VA；Metkinen Oy，Finland）。

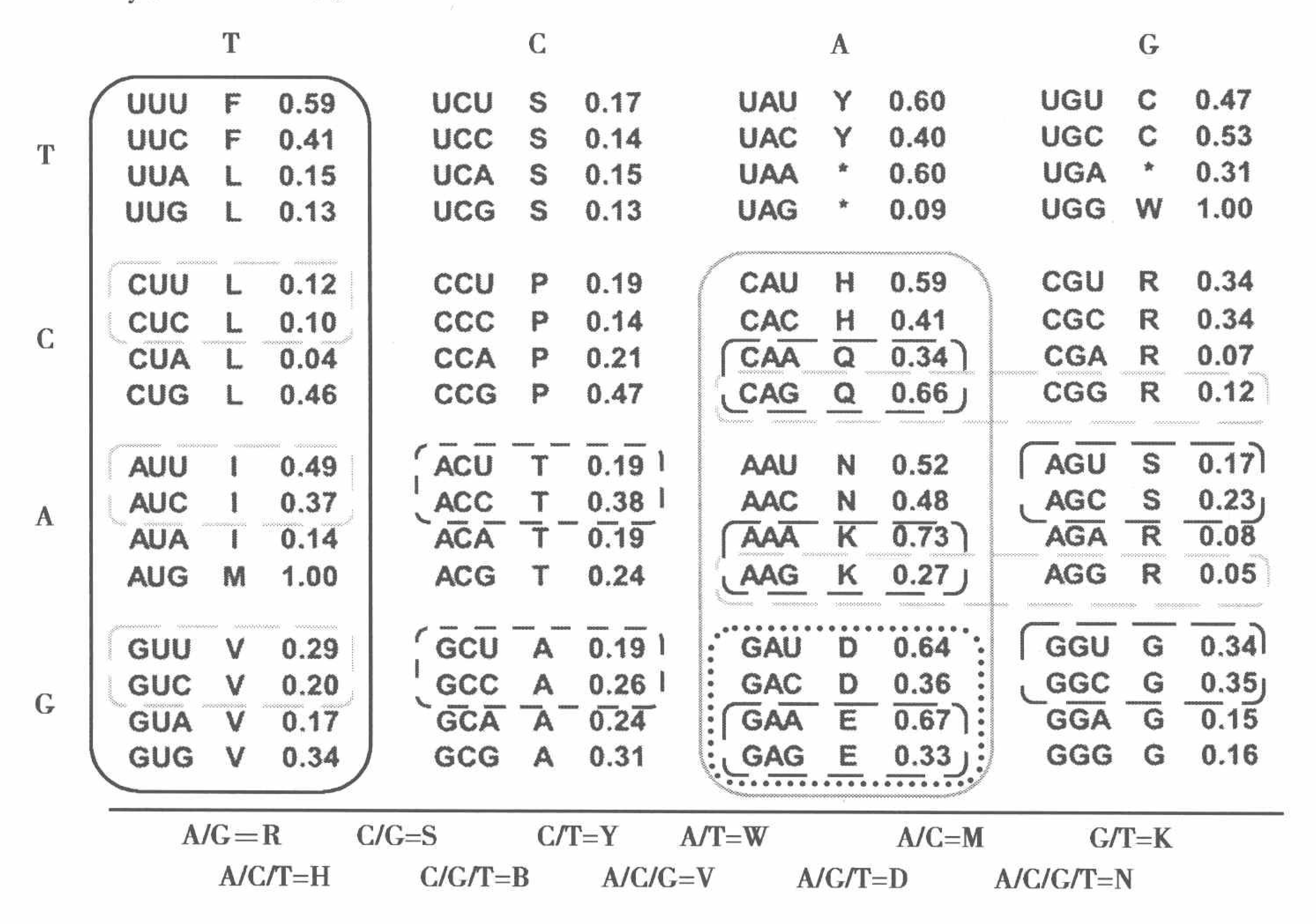


图 3.2 用于盘绕螺旋库设计的密码子及其可能的组合。显示出大肠杆菌的密码子及偏向：标准规范三联密码子，随后是密码子所代表的氨基酸和该有机体中该密码子所占 h 分数（http://www.kazusa.or.jp/codon；参考文献 [83]）。图中还显示了代表一组混合核苷酸的字母，这些字母随后会导致产生（盘绕螺旋）库所需的简并的碱基。方框给出盘绕螺旋设计所需的各种组合。

3.2.5.2　密码子偏向

密码子偏向是指一种而不是另外一种密码子将编码一个特定的氨基酸（图 3.2A）。例如，在大肠杆菌中，CGT 编码 Arg 的次数比 CGA 多 5 倍。很明显，当用包含编码库中多肽的基因的质粒转染细胞（如大肠杆菌）时，必须选用在宿主有机体中常用的三联体，而不是不常用的密码子，后者可能导致可读框移位[80～82]或库中代表性不好的序列。还有，在设计用于筛选的库的情况。例如，希望引进 β-分支氨基酸的所有组合到 a 位，但这取决于密码子的用法。图 3.2 显示某些我们最喜好的密码子组合，用于盘绕螺旋的不同位置。大多数组合只包括少数选项以使库维持在合理的大小。通过检查图 3.3（大肠杆菌密码子用法；对此或其他有机体见 www.kazusa.or.jp/condon）我们见到，如果要盘绕螺旋的 a 位包括 Val、Ile 和 Thr，需要密码子 GTN、ATH（如果 Met 必须避免）和 ACN。这会意味着在密码子的第一位需要 G 或 A，第二位需要 T 或 C，并且第三位需要除 G 以外的任何其他核苷酸，因为 ATG 会编码 Met，导致简并的密码子 RYH（所有作为结果的密码子都是大肠杆菌经常使用的）。但是，对于 Val、Ile 和 Thr 的组合，不可能排除库中不需要的氨基酸，因为上述核苷酸的混合也会包括编码 Ala 的 GCH。

密码子	NNK	NTN	NTK	DTK	VTY	RSY	VAN	VAR	GAN	MRG	
Gly	2	–	–	–	–	2	–	–	–	–	
Ala	2	–	–	–	–	2	–	–	–	–	
Val	2	4	2	2	2	–	–	–	–	–	
Leu	3	6	3	1	2	–	–	–	–	–	
Ile	1	3	1	1	2	–	–	–	–	–	脂肪族侧链
Met	1	1	1	1	–	–	–	–	–	–	
Pro	2	–	–	–	–	–	–	–	–	–	
Phe	1	2	1	1	–	–	–	–	–	–	
Trp	1	–	–	–	–	–	–	–	–	–	
Ser	3	–	–	–	–	2	–	–	–	–	
Thr	2	–	–	–	–	2	–	–	–	–	
Asn	1	–	–	–	–	–	2	–	–	–	
Gln	1	–	–	–	–	–	2	2	–	1	极性侧链
Tyr	1	–	–	–	–	–	–	–	–	–	
Cys	1	–	–	–	–	–	–	–	–	–	
Lys	1	–	–	–	–	–	2	2	–	1	
Arg	3	–	–	–	–	–	–	–	–	2	
His	1	–	–	–	–	–	2	–	–	–	带电侧链
Asp	1	–	–	–	–	–	2	–	2	–	
Glu	1	–	–	–	–	–	2	2	2	–	
stop (TAG)	1	–	–	–	–	–	–	–	–	–	
总计	32	16	8	6	3	4	6	3	2	4	

图 3.3　列出的是对应于图 3.2 所示密码子随机化框架的氨基酸频度。另外，也列出了 NNK 的分布。当包括所有 20 种氨基酸，且终止密码子须尽量排除时，这一组合得到广泛应用。"Total"给出在计算库容的时候要用的氨基酸数。NTN 或 VTY 编码的疏水氨基酸组合（图 3.2 中分别用黑线或灰而长的虚线围起来）。为尽量减少 Ile 和 Leu 的过表示，NTN 可以缩小为 NTK 或 DTK。不同的极性和带电组合分别由 VAN（灰线）、VAR（黑长虚线）、GAN（黑点线）或 MRG（灰短虚线）编码。如果期望的话，可能的链套区可由 RSY（黑短虚线）编码得到。请注意，在我们的例子中，第 3 碱基是混合的，以增加密码子使用中的变化，这可能对重复序列是重要的。但是，如果愿意的话，在多数情况下，最后一个碱基可保持不变。在 Gln、Arg 和 Lys 的情形，MRG 只含 Arg 的一个稀有密码子（AGG），而组合 MRR 含 Arg 的 3 个稀有密码子（CGG、AGA 和 AGG）这（保持第三个碱基不变）就特别重要。组合 NTN 和 VAN 先前在 Hecht 组发展起来的两元设计样式中使用过（见参考文献[84]和[85]，以及第 9 章）。

另一个例子是，如果想包括氨基酸 Gln、Asn 和 Lys，则根据默认的编码，必须也包括 His。但通过合成只编码那些需要的氨基酸的三联核苷酸，这个问题是可以克服的[31,79]。另外，必须仔细排出表示得不好的密码子。例如，在表示 Arg 时就会遇到这种情况，因为在大肠杆菌中短缺 AGG 和 AGA 密码子的氨酰基 tRNA 合成酶。最后必须注意，如果一个位点在库中需要无偏向的表示，那么在那个位点上对所有氨基酸都应该选用 1∶1 的比例（与代表它们的密码子一样）。在一个简并位上，某一个氨基酸比另一个氨基酸更多地表达，会自然地把对应于此差异的偏向引入体系中，因为那个氨基酸得到了更多的表达。

3.2.5.3 筛选系统

向盘绕螺旋开放的最常用筛选系统是蛋白质片段互补测试选择。在这个测试中，相互作用的蛋白质，如两个盘绕螺旋片段，被系在报告蛋白的两边上，仅当两个融合蛋白或多肽结合的时候报告蛋白才被激活。这已被用于二氢叶酸还原酶[8,31,70,86~88]、泛素[89]、β-半乳糖苷酶[90]、β-内酰胺酶[91,92]和绿色荧光蛋白[93]这样的胞内测试，有同时抗蛋白酶敏感或抗毒性多肽选择的额外优点。

用于盘绕螺旋的另一个筛选系统是酵母双杂交系统。在这个系统中，一个螺旋被融合于 Gal4p 转录因子的结合结构域（它与一个启动子结合），而另一个螺旋则融合于其活性结构域（它与聚合酶相互作用）。只有两个螺旋间的相互作用能够把两个嵌合蛋白拉到互相接近的位置并允许转录激活子起作用，并因而激活可以测试活性的报告基因（如 β-半乳糖苷酶）。使用这一系统，可从一蛋白质库中筛选到一个蛋白质或结构域的新的相互作用伙伴（最新的述评，见参考文献 [94]）。

λ 阻抑子系统，是基于通过用一杂寡聚化的结构域取代其 C 端结构域，使大肠杆菌 λ 阻抑子的活性重新恢复。当 C 端结构域与自身（同型相互作用）或来自另一个融合蛋白的不同的结构域（异型相互作用）形成二聚体（或更高的寡聚体）时，就探测到阻抑作用。或者通过模拟噬菌体感染免疫，或者通过检测对 β-半乳糖苷酶那样的报告子的阻抑，可以探测到功能阻抑子的存在[95,96]。

最后，通过噬菌体展示，蛋白质可以展示在丝状噬菌体的表面外壳蛋白上。从展示在噬菌体上的多肽库中可以挑选结合者通过几轮挑选来浓缩。这一技术也容许挑选与非天然化合物结合的蛋白质[97]。

3.2.5.4 与设计有关的计算

某些研究组不使用体内选择系统，而是使用计算（*in silico*）的方法得到盘绕螺旋的有希望的序列。本节讨论与此有关的问题。

（1）Keating 等发展了一种计算方法，以预测疏水核心突变对相互作用特异性的效应[98]。通过使用一种算法，综合范德华堆积作用能，溶剂化效应和 α 螺旋倾向性，来达到此目的。据此，可预测来自于核心堆积的伴随偏好。设计了核心 a 和 d 残基突变为 Leu、Ile 和 Val 的盘绕螺旋，以产生 6 个不同稳定性的不同的杂二聚体（见注 32）。该算法可以预测稳定性，并与实验数据很符合。

（2）Mayo 的小组使用设计的算法估算表面蛋白相互作用。他们检测了 3 个打分函数：氢键位能，与没有补偿的极性氢掩埋罚分相结合的氢键位能，氢键位能与螺旋倾

向结合[69]。对这 3 个打分函数中的每一个，以 GCN4-p1 为例（见注 2），用该算法找到优化的氨基酸序列，并合成相应的多肽。得到的所有多肽都是二聚的；在 1°C，近 100%是螺旋的；较之于具有随机分布的表面亲水残基的 GCN4 多肽的 15°C，熔融温度为 69～72°C。数据建议，在盘绕螺旋多肽表面残基的序列设计中，螺旋倾向性是关键因素。

(3) 为了要生成具有右手超螺旋缠绕的盘绕螺旋，发展出了一种算法，即加入了对主链游变性的处理[99]。因为要用固定主链的话，一个已经存在的例子便可提供坐标定位，即使这样，这些数值对近同源调整后的序列可能并不适合。通过骨架的游变性，可从一小组主链构象中取样。这些取样然后再与侧链堆积取样相耦合。结果，以计算方法设计出了右手二聚的、三聚的以及四聚的盘绕螺旋（见注 33）。疏水极性残基构成的样式规定蛋白质的整体折叠。寡聚态、主链构象以及侧链旋转异构通过计算，在不同的主链结构中找到最好的堆积来选择。得到的多肽正确地形成与设计目标一致的寡聚态，四聚体 X 射线结构与设计的结构在原子细节上相符。

(4) Harbury 的一篇文章结合了正向（期望的结构）和负向（远离不期望的其他结构）设计，以优化相互作用的特异性［见 3.2.2.2 (3)；参考文献［39］］。GCN4 中心七元重复的 a 位、d 位、e 位和 g 位改变为所有非脯氨酸残基，以产生大约 8×10^9 个可能的序列（见注 15)。该算法类似于计算的双突变循环，其中的多肽是序列优化的，而不是结构优化的。考虑过的能量竞争态有：

a. 折叠齐二聚期望结构；

b. 能量上有利的杂二聚体；

c. 未折叠的能量不稳态；

d. 难溶聚集态。

对于候选序列，这 4 种竞争态中的每一种都计算了自由能。这种所谓的“多态”设计比单态设计体系有一个优点，后者只寻找目标的最低自由能。不像这些设计，多态设计的结构常常偏离 PV 假定，并且同时，考虑到一些因素，如引起聚集的过多的疏水暴露和降低稳定性的过多的极性掩埋。更适当地，这种设计考虑了所有这些因素，当这些竞争的力支配对目标态的选择时，这种设计才需要改变。因而，选择是稳定性和特异性间平衡的结果，而不仅仅是目标稳定性。

3.3 注

1. GCN4-p1 的序列：Ac-R MKQLEDK VEELLSK NYHLENE VARLKKL VGER-COOH。在 Harbury 等的研究中[16,17]突变过的 **a** 和 **d** 残基用黑体显示。经受过许多突变研究的核心 Asn 16 标有下画线。

2. 设计的杂二聚肽拉链[9]为合成多肽：Acid-p1（Ac-AQLEKE LQALEKE NAQLEWE LQALEKE LAQ-NH_2）和 Base-p1（Ac-AQLKKK LQALKKK NAQLK-WK LQALKKK LAQ-NH_2）。在此和随后的注中，序列中单个七元重复用空格分开。核心 Asn 残基，如已说过的经受过突变，见 3.2.1.2 (4)，标有下画线。

3. 两个半胱氨酸二硫键联结的多肽的序列 2H（Ac-KCEALEGK LEALEGK LEAAEGK LEALEGK LEALEG-NH_2 和 Ac-ELAELKGE LAELKGE AAELKGE

LAELKGE LAECKG-NH2）和 4H（Ac-KCEALEGK LEALEGK LEAAEGK LEALEGK LEALEG-NH2 和 Ac-ELAELKGE LAELKGE LAEAKGE LAELKGE LAECKG-NH2）。为显示 2H 和 4H 中两条链间的共价连接点，半胱氨酸用黑体示出，而规定寡聚态的 Ala 用下画线示出[20]。

4. 鼠 COMP 蛋白的盘绕螺旋结构域（氨基酸 27～72）序列（21）：GDL APQMLRE LQETNAA LQDVREL LRQQVKE ITFLKNT VMECDAC G。在鼠 COMP 的表达片段中，Gly 27 被 Met 所取代。

5. 研究基于多肽 A1（MRGSHHHHHHGSMA SGDLENE YAQLERE VRSLEDE AAELEQK VSRLKNE IEDLAEI GDLNNTSGIRRPAA KLN）。三氟亮氨酸和六氟亮氨酸的掺入以取代亮氨酸（黑体），通过在没有亮氨酸的培养液中加入三氟亮氨酸或六氟亮氨酸，用基因表达[22,23]。

6. 如注 2 所说明的，Kretsinger 等的研究系基于 GCN4-p1 多肽。他们将 C 端酰胺化，并报道了在第一个七元重复的 Asp 和 Lys 间插入 Ser 的序列[24]。因为这一插入会移动七元重复，我们可以假定这是图中的一个错误。有下画线的 Asn（见注 2）被变为 Asp 二氨基庚二酸（diaminopropionic acid）及其一、二和三甲基化类似物。

7. 本节中，杂二聚体系从两个设计的盘绕螺旋库中选出：LibA：VAQL＃E＃ VKTL＃A＃ §YEL＃S＃ VQRL＃E＃ VAQL 和 LibB：VDEL＃A＃ VDQL＃D＃ §YAL＃T＃ VAQL＃K＃ VEKL，其中，＃表示 E、Q、K 和 R 的等摩尔混合，§表示 V 和 N 的等摩尔混合[31]。核心**a** 位和**d** 位的序列来自于 GCN4（见注 2），而**b** 位、**c** 位和 **f** 位的序列（有下画线）来自 c-Jun（IARLEEK VKTLKAQ NYELAST ANMLREQ VAQL）和 c-Fos（TDTLQAE TDQLEDE KYALQTE IANLLKE KEKL）的盘绕螺旋结构域。

8. GCN4-pVL 变体的序列是 Ac-R MKQLEDK VEE＃LSK §YHLENE VARLKKL VGER，其中**a** 位和**d** 位用黑体显示。用＃表示的 12 位（**d**）或者是 Leu，或者是一个极性残基（N、Q、S 或 T），用§表示的 16 位（**a**）或者是 Val，或者是一个极性残基[25]。

9. 两项研究都使用二硫键桥接的盘绕螺旋。此螺旋基于序列 VGALKKE，做了一些修正以避免链内和链间与替换位点（**X**）的电荷-电荷相互作用，并调节整体电荷。其序列分别是 Ac-CGGE VGALKAQ VGALQAQ XGALQKE VGALKKE VGALKK-NH2[33] 和 Ac-CGGE VGALKAE VGALKAQ IGA**X**-QKQ IGALQKE VGALKK-NH2[32]。

10. Ji 等使用了记为 N36（L6）C34 的猿猴免疫缺陷病毒 gp41 核心的重组模型，其中的氨基端螺旋（N36）构成中心的三聚盘绕螺旋，而羧基端螺旋（C34）以反平行的取向堆积到此盘绕螺旋三聚体的疏水凹槽中。N36 和 C34 被一短联结子分开（L6；参考文献［34］）。在 N36 结构域中，极性核心残基（黑体）突变为 Ile：AGIVQQ QQQLLDV VKRQQEL LRLTVWG TKNLQTR VT。产生不溶聚集的 Q→I 突变用下画线标出。

11. 父辈多肽 Lac21 的序列是 Ac-MKQLADS LMQLARQ VSRLESA-NH2（见注 28）。有下画线的残基被突变为 E 或 K，以分别产生形成杂四聚的多肽 Lac21E 和 Lac21K[35]。

12. APC-55 的序列：AAAS YDQLLKQ VEALKME NSNLRQE LEDNSNH LT-

KLETE ASNMKEV LKQLQGS I 和 anti-APCp1 的序列：MAAK GDQLKKE VEALEYE NSNLRKK LEDHKKK LTKLKTE ISNAKKM LKQLYAS I[36]。anti-APCp1 与 APC-55 相比较的核心变化用黑体标出；e 位和 g 位的改变用下画线标示；为增加稳定性、增加净电荷以便于纯化，以及为加入发色团的改变标为斜体。

13. 三条肽链是 T_9：Ac-R MKQLEKK **X**EELLSK **A**QQLEKE **A**AQLKKL VG-NH_2，T_{16}：Ac-R MKQLEKK **A**EELLSK **X**QQLEKE **A**AQLKKL VG-NH_2，T_{23}：Ac-R MKQLEKK **A**EELLSK **A**QQLEKE **X**AQLKKL VG-NH_2[37]。在所有 3 个多肽中不相同的残基标为黑体，**X** 表示环己基丙氨酸残基。

14. 序列基于 Acid-pLL 和 Base-pLL，它们与 Acid-p1 和 Base-p1（见注 1）相同，但核心 Asn 突变为 Leu（也见 3.2.1.2 小节；第 4 条）。对 Base-pLL 的两个 L→K 突变（黑体）得到 Base-pK：Ac-AQLKKK LQALKKK **K**AQLKWK **K**QALKKK LAQ-NH_2[38]。

15. 研究基于 GCN4 的 N 端戴帽变体（见注 31；参考文献 [7]），Asn 16 移动一个七元重复到第 9 位，产生多肽 p-CAP：**S** VKELEDK **N**EELLS**X** **X**YH**XX**NE VARLKKL VGER。与 GCN4-p1 的变化（见注 2）用黑体标出。**X** 表示在设计计算中允许改变的位置[39]。

16. 研究基于设计的齐二聚盘绕螺旋 EK：Ac-K CGALEKK LGALEKK AGALEKK LGALEKK LGALEK-NH_2。制造了 3 个盘绕螺旋突变体，其中

a. EK 中 5 个 e 位 Glu 残基（有下画线）被突变为 Gln 残基（QK 肽）；

b. 5 个 g 位 Lys 残基（有下画线）被突变为 Gln 残基（EQ 肽）；

c. 上两个变化的合并（QQ 肽）。

用双突变循环分析，计算了螺旋间离子吸引对盘绕螺旋稳定性的能量贡献[41]。

17. 父辈的卵黄原生成素结合蛋白 ER_{34} 的序列：ITIR AAFLEKE NTALRTE VAELRKE VGRCRNI VSKYETR YGPL。有下画线的 e 和 g 残基在后来的多肽中被改变[45]。

18. 选择出的最成功的 WinZip-A2B1（见注 7）由多肽 WinZip-A2（Ac-STT VAQLRER VKTLRAQ NYELESE VQRLREQ VAQL AS-NH_2）和 WinZip-B1（Ac-STS VDELQAE VDQLQDE NYALKTK VAQLRKK VEKL SE-NH_2）组成[8]。

19. 研究基于设计的 E-和 K-肽。E-肽：Ac-E LGALEKE LGALEKE **L**GALEKE LGALEKE LGALEK-NH_2；K-肽：Ac-K LGALKEK LGALKEK **L**GALKEK LGALKEK LGALKE-NH_2。16 位和 19 位（黑体）被变为 Ala 以产生不同的 Leu-Ala 核心组合，2 位或 33 位（有下画线）被变为 Cys 以允许在平行或反平行态下有二硫桥[52]。

20. APH 的序列：MKQLEKE LKQLEKE LQA**I**EKQ LAQLQWK **A**QA*R*KKK LAQLKKK LQA[55]。空间匹配的核心残基（Ile 和 Ala）用黑体显示；设计的 N 端谷酰胺和 C 端赖氨酸的库仑相互作用用下画线标出。且单 Arg 残基被放在 d 位（斜体），以启动二聚体生成。

21. 得到的多肽称为 Acid-a1（Ac-AQLEKE LQALEKE **LA**QLEWE **N**QALEKE LAQ-NH_2）和 Base-a1（Ac-AQLKKK LQA**N**KKK **LA**QLKWK LQALKKK LAQ-NH_2）[56]。相对于 Acid-p1 和 Base-p1（见注 1）的改变用黑体标出。

22. 多肽 Acid-RdL（Ac-AQLEKE LQALEKE LAQ**R**EWE LQALEKE LAQ-

NH_2）和 Base-EgL（Ac-AQLKKK LQALKKE LAQLKWK LQALKKK LAQ-NH_2）是基于 Acid-a1 和 Base-a1（见注 21），设计的极性相互作用用黑体[59]。

23. 合成了 5 个带有 N 或 C 端 Cys 和核心 Ala 残基的不同的多肽。其中 3 个多肽（C2A16、C33A16 和 C33A19）是基于七元重复 LEALEGK：Ac-K LEALEGK LEALEGK LEALEGK LEALEGK LEALEG-NH_2，并把 Cys 或者放在 2 位或者放在 33 位（下画线标明），Ala 或者放在 16 位或者放在 19 位（黑体）。另外两个多肽（C33A16 和 C33A19）则基于七元重复 LAELKGE：Ac-E LAELKGE LAELKGE LAELKGE LAELKGE LAECKG-NH_2，以 Cys 放在 33 位，而 Ala 或者放在 16 或者放在 19 为[49]。

24. 多肽 Acid-Kg（Ac-AQLEKE LQALEKK LAQLEWE NQALEKE LAQ-NH_2）是基于 Acid-a1（见注 21），而多肽 Base-Eg（Ac-AQLKKK LQANKKE LAQLKWK LQALKKK LAQ-NH_2）是基于 Base-a1。变化用黑体标出。

25. 合成了具有下列序列的多肽：Ac-（K LEALEG)$_n$-K-NH_2，其中 $n=1\sim5$[62]，并与羧酰胺甲基化的 α-原肌球蛋白在 190 位的半胱氨酸（CM-原肌球蛋白）比较。

26. 设计了一系列包含 9 个、12 个、16 个、19 个、23 个、26 个、30 个、33 个和 35 个氨基酸残基的多肽。其中 35 肽具有序列：Ac-E iealkae iealkae iealkae iealkae ieacka-NH_2[63]。较短的肽分别含有从 C 端算起的对应数量的氨基酸。

27. 这项研究使用肽 Succ-DELERR IRELEAR IK-NH_2[64]。Succ 指琥珀酰化的 N 端。

28. 研究过的多肽是 Lac 21：Ac-MKQLADS LMQLARQ VSRLESA-NH_2，Lac 28：Ac-LMQLARQ MKQLADS LMQLARQ VSRLESA-NH_2 和 Lac 35：Ac-LMQLARQ LMQLARQ MKQLADS LMQLARQ VSRLESA-NH_2[65]。

29. 研究系基于由具有序列 Ac-（E＃§ALEK)$_n$-NH_2 的 E-肽和具有序列 Ac-（K＃§ALKE)$_n$-NH_2 的 K-肽组成的 E/K 杂二聚体。此处 $n=3$ 或 4；＃指 I 或 V；§指 A 或 S[67]。

30. 多肽序列是 Ac-Q CGALQKQ VGALQKQ VGALQKQ VGALQKQ VGALQK-NH_2。标了下画线的位置 1、6、15、20 和 34 被突变为 Gln[73]。

31. 此项研究使用重组的 GCN4-pMSE 肽（MS VKELEDK VEELLSK NYHLENE VARLKKL VGER)。戴帽模体用黑体标示，与 GCN4-p1（见注 2）相比较的突变用下画线标明。本研究中的其他多肽是 GCN4-pSE，它缺少起始的甲硫氨酸和 GCN4-pAA，其中 GCN4-pSE 的 Ser 和 Glu 被突变为 Ala[77]。GCN4-pAA 的稳定性与 GCN4-p1 可比，而 GCN4-pSE 和 GCN4-pMSE 变体比 GCN4-pAA 分别稳定了 0.5kcal/mol 和 1.2kcal/mol。因而，末端 Met 残基的疏水贡献是 0.7kcal/mol。

32. 为 GCN4（见注 2）Aicd/Base 杂二聚体而设计的杂二聚盘绕螺旋 GABH 有酸性序列 A（Ac-E VKQLEAE VEE＃ESE ＃WHLENE VARLEKE NAECEA-NH_2）和碱性序列 B（Ac-K VKQLKAK VEE＃KSK ＃WHLKNK VARLKKK NAECKA-NH_2)[98]。位置 **d** 12 和 **a** 16（＃）被分别突变为 Val、Ile 和 Leu，以生成多肽 A_{LL}、A_{IL}、A_{LI}、B_{LL} 和 B_{LV}。

33. 设计的序列是二聚的 RH2（Ac-AE IEQLKKE §AYL IKKLKAEKLAE IKKLKQEKA-NH_2)、三聚的 RH3（Ac-AE ＃EQ＃KKEIAYL ＃KK＃KAEILAE ＃KK＃KQEIA-NH_2）和四聚的 RH4（Ac-AE LEQ＃KKEIAYL LKK＃KAEIL AE

LKK#KQEIA-NH_2)[99]。十一元重复（**a**～**k**）的疏水残基（**a**、**d**和**h**）用黑体标明；§标示正缬氨酸；#标示别异亮氨酸残基。

致谢 本工作得到德国 Forschungsgemeinschaft 的 Emmy Noether 项目的支持（grant Ar 373）。

参考文献

1. Wolf, E., Kim, P. S., and Berger, B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* 6, 1179-1189.
2. Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.* 21, 375-382.
3. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 240, 1759-1764.
4. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* 11, 82-88.
5. Kohn, W. D., Mant, C. T., and Hodges, R. S. (1997) Alpha-helical protein assembly motifs. *J. Biol. Chem.* 272, 2583-2586.
6. O' Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254, 539-544.
7. DeLano, W. L. (2002) The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA; http://www.pymol.org.
8. Arndt, K. M., Pelletier, J. N., Müller, K. M., Plückthun, A., and Alber, T. (2002) Comparison of *in vivo* selection and rational design of heterodimeric coiled coils. *Structure* 10, 1235-1248.
9. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) Peptide 'Velcro': design of a heterodimeric coiled coil. *Curr. Biol.* 3, 658-667.
10. Mason, J. M. and Arndt, K. M. (2004) Coiled coil domains: stability, specificity, and biological implications. *Chem. Biochem.* 5, 170-176.
11. Müller, K. M., Arndt, K. M., and Alber, T. (2000) Protein fusions to coiled-coil domains. *Methods Enzymol.* 328, 261-282.
12. Arndt, K. M., Müller, K. M., and Plückthun, A. (2001) Helix-stabilized Fv (hsFv) antibody fragments: substituting the constant domains of a Fab fragment for a heterodimeric coiled-coil domain. *J. Mol. Biol.* 312, 221-228.
13. Pack, P., Müller, K. M., Zahn, R., and Plückthun, A. (1995) Tetravalent miniantibodies with high avidity assembling in *Escherichia coli*. *J. Mol. Biol.* 246, 28-34.
14. Naik, R. R., Kirkpatrick, S. M., and Stone, M. O. (2001) The thermostability of an alpha-helical coiled-coil protein and its potential use in sensor applications. *Biosens. Bioelectron.* 16, 1051-1057.
15. Crick, F. H. S. (1953) The packing of α-helices: simple Coiled Coils. *Acta Crystallogr.* 6, 689-697.
16. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262, 1401-1407.
17. Harbury, P. B., Kim, P. S., and Alber, T. (1994) Crystal structure of an isoleucinezipper trimer. *Nature* 371, 80-83.
18. Betz, S. F., Bryson, J. W., and DeGrado, W. F. (1995) Native-like and structurally characterized designed alpha-helical bundles. *Curr. Opin. Struct. Biol.* 5, 457-463.
19. Woolfson, D. N. and Alber, T. (1995) Predicting oligomerization states of coiled coils. *Protein Sci.* 4, 1596-1607.
20. Monera, O. D., Sonnichsen, F. D., Hicks, L., Kay, C. M., and Hodges, R. S. (1996) The relative positions of alanine residues in the hydrophobic core control the formation of two-stranded or four-stranded alpha-helical coiled-coils. *Protein Eng.* 9, 353-363.

21. Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T., and Engel, J. (1996) The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel? *Science* 274, 761-765.

22. Tang, Y. and Tirrell, D. A. (2001) Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. *J. Am. Chem. Soc.* 123, 11089-11090.

23. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) Fluorinated coiled-coil proteins prepared *in vivo* display enhanced thermal and chemical stability. *Angew. Chem. Int. Ed.* 40, 1494-1496.

24. Kretsinger, J. K. and Schneider, J. P. (2003) Design and application of basic amino acids displaying enhanced hydrophobicity. *J. Am. Chem. Soc.* 125, 7907-7913.

25. Akey, D. L., Malashkevich, V. N., and Kim, P. S. (2001) Buried polar residues in coiled-coil interfaces. *Biochemistry* 40, 6352--6360.

26. Glover, J. N. and Harrison, S. C. (1995) Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* 373, 257-261.

27. Gonzalez, L., Jr., Woolfson, D. N., and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol.* 3, 1011-1018.

28. Junius, F. K., Mackay, J. P., Bubb, W. A., Jensen, S. A., Weiss, A. S., and King, G. F. (1995) Nuclear magnetic resonance characterization of the Jun leucine zipper domain: unusual properties of coiled-coil interfacial polar residues. *Biochemistry* 34, 6164--6174.

29. Potekhin, S. A., Medvedkin, V. N., Kashparov, I. A., and Venyaminov, S. (1994) Synthesis and properties of the peptide corresponding to the mutant form of the leucine zipper of the transcriptional activator GCN4 from yeast. *Protein Eng.* 7, 1097-1101.

30. Lumb, K. J. and Kim, P. S. (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* 34, 8642-8648.

31. Arndt, K. M., Pelletier, J. N., Müller, K. M., Alber, T., Michnick, S. W., and Plückthun, A. (2000) A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. *J. Mol. Biol.* 295, 627-639.

32. Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T., and Hodges, R. S. (2000) Effects of side-chain characteristics on stability and oligomerization state of a *de novo*-designed model coiled-coil: 20 amino acid substitutions in position "d". *J. Mol. Biol.* 300, 377-402.

33. Wagschal, K., Tripet, B., Lavigne, P., Mant, C., and Hodges, R. S. (1999) The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. *Protein Sci.* 8, 2312-2329.

34. Ji, H., Bracken, C., and Lu, M. (2000) Buried polar interactions and conformational stability in the simian immunodeficiency virus (SIV) gp41 core. *Biochemistry* 39, 676-685.

35. Fairman, R., Chao, H. G., Lavoie, T. B., Villafranca, J. J., Matsueda, G. R., and Novotny, J. (1996) Design of heterotetrameric coiled coils: evidence for increased stabilization by Glu (—) -Lys (+) ion pair interactions. *Biochemistry* 35, 2824-2829.

36. Sharma, V. A., Logan, J., King, D. S., White, R., and Alber, T. (1998) Sequencebased design of a peptide probe for the APC tumor suppressor protein. *Curr. Biol.* 8, 823-830.

37. Schnarr, N. A. and Kennan, A. J. (2002) Peptide tic-tac-toe: heterotrimeric coiledcoil specificity from steric matching of multiple hydrophobic side chains. *J. Am. Chem. Soc.* 124, 9779-9783.

38. Campbell, K. M. and Lumb, K. J. (2002) Complementation of buried lysine and surface polar residues in a designed heterodimeric coiled coil. *Biochemistry* 41, 7169-7175.

39. Havranek, J. J. and Harbury, P. B. (2003) Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* 10, 45-52.

40. Kohn, W. D., Kay, C. M., and Hodges, R. S. (1995) Protein destabilization by electrostatic repulsions in the two-stranded alpha-helical coiled-coil/leucine zipper. *Protein Sci.* 4, 237-250.

41. Zhou, N. E., Kay, C. M., and Hodges, R. S. (1994) The net energetic contribution of interhelical electrostatic attractions to coiled-coil stability. *Protein Eng.* 7, 1365-1372.

42. Graddis, T. J., Myszka, D. G., and Chaiken, I. M. (1993) Controlled formation of model homo- and heterodimer coiled coil polypeptides. *Biochemistry* 32, 12664-12671.

43. Moll, J. R., Olive, M., and Vinson, C. (2000) Attractive interhelical electrostatic interactions in the proline- and acidic-rich region (PAR) leucine zipper subfamily preclude heterodimerization with other basic leucine zipper subfamilies. *J. Biol. Chem.* 275, 34826-34832.

44. Krylov, D., Mikhailenko, I., and Vinson, C. (1994) A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J.* 13, 2849-2861.

45. Krylov, D., Barchi, J., and Vinson, C. (1998) Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J. Mol. Biol.* 279, 959-972.

46. Newman, J. R. and Keating, A. E. (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* 300, 2097-2101.

47. Oakley, M. G. and Hollenbeck, J. J. (2001) The design of antiparallel coiled coils. *Curr. Opin. Struct. Biol.* 11, 450-457.

48. Monera, O. D., Zhou, N. E., Kay, C. M., and Hodges, R. S. (1993) Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization. *J. Biol. Chem.* 268, 19218-19227.

49. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two stranded alpha-helical coiled-coils. *Biochemistry* 33, 3862-3871.

50. Oakley, M. G. and Kim, P. S. (1997) Protein dissection of the antiparallel coiled coil from *Escherichia coli* seryl tRNA synthetase. *Biochemistry* 36, 2544-2549.

51. Kohn, W. D. and Hodges, R. S. (1998) *De novo* design of α-helical coiled coils and bundles: models for the development of protein-design principles. *Trends Biotechnol.* 16, 379-389.

52. Monera, O. D., Zhou, N. E., Lavigne, P., Kay, C. M., and Hodges, R. S. (1996) Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *J. Biol. Chem.* 271, 3995-4001.

53. Holton, J. and Alber, T. (2004) Automated protein crystal structure determination using ELVES. *Proc. Natl. Acad. Sci. USA* 101, 1537-1542.

54. Gonzalez, L., Jr., Plecs, J. J., and Alber, T. (1996) An engineered allosteric switch in leucine-zipper oligomerization. *Nat. Struct. Biol.* 3, 510-515.

55. Gurnon, D. G., Whitaker, J. A., and Oakley, M. G. (2003) Design and characterization of a homodimeric antiparallel coiled coil. *J. Am. Chem. Soc.* 125, 7518-7519.

56. Oakley, M. G. and Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* 37, 12603-12610.

57. Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., and Leberman, R. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 A. *Nature* 347, 249-255.

58. Stebbins, C. E., Borukhov, S., Orlova, M., Polyakov, A., Goldfarb, A., and Darst, S. A. (1995) Crystal structure of the GreA transcript cleavage factor from *Escherichia coli*. *Nature* 373, 636-640.

59. McClain, D. L., Gurnon, D. G., and Oakley, M. G. (2002) Importance of potential interhelical salt-bridges involving interior residues for coiled-coil stability and quaternary structure. *J. Mol. Biol.* 324, 257-270.

60. McClain, D. L., Woods, H. L., and Oakley, M. G. (2001) Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation. *J. Am. Chem. Soc.* 123, 3151-3152.

61. Litowski, J. R. and Hodges, R. S. (2001) Designing heterodimeric two-stranded alpha-helical coiled-coils: the effect of chain length on protein folding, stability and specificity. *J. Pept. Res.* 58, 477-492.

62. Lau, S. Y., Taneja, A. K., and Hodges, R. S. (1984) Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-

coils. *J. Biol. Chem.* 259, 13253-13261.

63. Su, J. Y., Hodges, R. S., and Kay, C. M. (1994) Effect of chain length on the formation and stability of synthetic alpha-helical coiled coils. *Biochemistry* 33, 15501-15510.

64. Burkhard, P., Meier, M., and Lustig, A. (2000) Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci.* 9, 2294-2301.

65. Fairman, R., Chao, H. G., Mueller, L., Lavoie, T. B., Shen, L., Novotny, J., and Matsueda, G. R. (1995) Characterization of a new four-chain coiled-coil: influence of chain length on stability. *Protein Sci.* 4, 1457-1469.

66. Kwok, S. C. and Hodges, R. S. (2004) Stabilizing and destabilizing clusters in the hydrophobic core of long two-stranded α-helical coiled-coils. *J. Biol. Chem.* 279, 21576-21588.

67. Litowski, J. R. and Hodges, R. S. (2002) Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. *J. Biol. Chem.* 277, 37272-37279.

68. O'Neil, K. T. and DeGrado, W. F. (1990) A thermodynamic scale for the helixforming tendencies of the commonly occurring amino acids. *Science* 250, 646-651.

69. Dahiyat, B. I., Gordon, D. B., and Mayo, S. L. (1997) Automated design of the surface positions of protein helices. *Protein Sci.* 6, 1333-1337.

70. Mason, J. M., Schmitz, M. A., Müller, K. M., and Arndt, K. M. (2006) Semirational design of Jon-Fos coiled coils with increased affinity: universal implications for leucine zipper prediction and design. *Proc. Natl. Acad. Sci. USA*, in press.

71. Richardson, J. S. and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices. *Science* 240, 1648-1652.

72. Dasgupta, S. and Bell, J. A. (1993) Design of helix ends. Amino acid preferences, hydrogen bonding and electrostatic interactions. *Int. J. Pept. Protein Res.* 41, 499-511.

73. Kohn, W. D., Kay, C. M., and Hodges, R. S. (1997) Positional dependence of the effects of negatively charged Glu side chains on the stability of two-stranded alphahelical coiled-coils. *J. Pept. Sci.* 3, 209-223.

74. Doig, A. J. (2002) Recent advances in helix-coil theory. *Biophys. Chem.* 101-102, 281-293.

75. Kumar, S. and Bansal, M. (1998) Dissecting alpha-helices: position-specific analysis of alpha-helices in globular proteins. *Proteins* 31, 460-476.

76. Aurora, R. and Rose, G. D. (1998) Helix capping. *Protein Sci.* 7, 21-38.

77. Lu, M., Shu, W., Ji, H., Spek, E., Wang, L., and Kallenbach, N. R. (1999) Helix capping in the GCN4 leucine zipper. *J. Mol. Biol.* 288, 743-752.

78. Sober, H. A. (1977) *CRC Handbook of Biochemistry and Molecular Biology*. 3rd ed., The Chemical Rubber Co, Cleveland, OH.

79. Virnekäs, B., Ge, L., Plückthun, A., Schneider, K. C., Wellnhofer, G., and Moroney, S. E. (1994) Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.* 22, 5600-5607.

80. Spanjaard, R. A. and van Duin, J. (1988) Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proc. Natl. Acad. Sci. USA* 85, 7967-7971.

81. Jung, S., Arndt, K. M., Muller, K. M., and Pluckthun, A. (1999) Selectively infective phage (SIP) technology: scope and limitations. *J. Immunol. Methods* 231, 93-104.

82. Arndt, K. M., Jung, S., Krebber, C., and Pluckthun, A. (2000) Selectively infective phage technology. *Methods Enzymol.* 328, 364-388.

83. Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28, 292; http://www.kazusa.or.jp/codon.

84. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262, 1680-1685.

85. West, M. W. and Hecht, M. H. (1995) Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Sci.* 4, 2032-2039.

86. Pelletier, J. N., Campbell-Valois, F X., and Michnick, S. W. (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. USA* 95, 12141-12146.

87. Pelletier, J. N., Arndt, K. M., Plückthun, A., and Michnick, S. W. (1999) An *in vivo* library-versus-library selection of optimized protein-protein interactions. *Nat. Biotechnol.* 17, 683-690.

88. Arndt, K. M., Jouaux, E. M., and Willemsen, T. (2004) Der richtige Dreh—Coiled Coils auf dem Weg zur Anwendung. *BioForum* 10, 48-49.

89. Johnsson, N. and Varshavsky, A. (1994) Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. USA* 91, 10340-10344.

90. Rossi, F., Charlton, C. A., and Blau, H. M. (1997) Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc. Natl. Acad. Sci. USA* 94, 8405-8410.

91. Wehrman, T., Kleaveland, B., Her, J. H., Balint, R. F., and Blau, H. M. (2002) Protein-protein interactions monitored in mammalian cells via complementation of beta-lactamase enzyme fragments. *Proc. Natl. Acad. Sci. USA* 99, 3469-3474.

92. Galarneau, A., Primeau, M., Trudeau, L. E., and Michnick, S. W. (2002) Betalactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein protein interactions. *Nat. Biotechnol.* 20, 619-622.

93. Ghosh, I., Hamilton, A. D., and Regan, L. (2000) Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* 122, 5658-5659.

94. Miller, J. and Stagljar, I. (2004) Using the yeast two-hybrid system to identify interacting proteins. *Methods Mol. Biol.* 261, 247-262.

95. Marino-Ramirez, L., Campbell, L., and Hu, J. C. (2003) Screening peptide/protein libraries fused to the lambda repressor DNA-binding domain in *E. coli* cells. *Methods Mol. Biol.* 205, 235-250.

96. Hu, J. C., O'Shea, E. K., Kim, P. S., and Sauer, R. T. (1990) Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science* 250, 1400-1403.

97. Willats, W. G. (2002) Phage display: practicalities and prospects. *Plant Mol. Biol.* 50, 837-854.

98. Keating, A. E., Malashkevich, V. N., Tidor, B., and Kim, P. S. (2001) Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proc. Natl. Acad. Sci. USA* 98, 14825-14830.

99. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom. *Science* 282, 1462-1467.

4 基于钙调素与荧光蛋白融合的钙指示剂

Kevin Truong, Asako Sawano, Atsushi Miyawaki, and Mitsuhiko Ikura

概要 钙调素（CaM）是普遍分布的蛋白质，参与钙介入的信号转导。当 Ca^{2+} 流入时，CaM 获得强亲和力，结合各种带有一个或多个 CaM 识别序列的胞内蛋白，启动或终止 Ca^{2+} 调控的信号串联。通过对 Ca^{2+}-CaM 复合物的核磁共振和晶体学结构研究，我们已得到对 CaM 目标识别机制的深入了解。一个最直接的应用就是，基于蛋白质的 Ca^{2+} 传感器。它使用了 CaM 复合物和绿色荧光蛋白，原先称为"chameleon"（钙变色子，英文原意为变色龙——译者）。钙变色子的主要优点是，它们可以在单个细胞中表达，并以特定组织或细胞为目标，测量局部 Ca^{2+} 的变化。本章描述有关的方法，包括钙变色子的克隆、鉴定它们的生物化学和生物物理特征，以及用数字荧光显微镜在单个细胞中使它们成像。
关键词 荧光共振能量转移（FRET）；钙调素（CaM）；绿色荧光蛋白（GFP）；钙变色子（chameleon）；钙信号传导

4.1 介　　绍

Ca^{2+} 浓度（$[Ca^{2+}]$）变化涉及细胞过程，如细胞发育、分化和凋亡，在它们存在的时空环境中研究这些变化，可以深入地了解它们的生物学意义。合成染料（如 Fura 和 Indo）和发光蛋白（aequorin）是研究 $[Ca^{2+}]$ 变化的常用工具，但是，多数合成染料迅速从细胞中流失，发光蛋白只有微弱的生物光且不能按比例显示 $[Ca^{2+}]$ 的变化[1]。相反，蛋白质 Ca^{2+} 传感器基于 Ca^{2+}-钙调素复合物和绿色荧光蛋白（起先被命名为钙变色子）克服了原先的这两个限制（见注 1）。本章我们描述包括钙变色子克隆、生化特性的鉴定和用数字荧光显微镜在单个细胞中成像的有关方法。在我们能够描述这些方法之前，必须理解荧光共振能量转移理论（FRET；见 4.1.1）和对钙变色子的一般设计（见 4.1.2）。

4.1.1 FRET（荧光共振能量转移）的概念

FRET 是能量从给体荧光团到受体荧光团的转移。如果给体的发射谱与受体的激发谱严重重叠，这种转移就能够发生。这种能量转移在距离小于 80Å 时发生，并当两个荧光团靠得最近且取向平行时效率最高[2,3]。在设计钙变色子时，用 GFP 突变体作为两个伙伴荧光团：蓝绿色荧光蛋白作为给体，而黄色荧光蛋白作为受体（见图 4.1；参考文献［4］）。如果两个荧光团融合于不同的作用伙伴，分子间 FRET 可以用来探测蛋白质-蛋白质相互作用，而如果两个荧光团都融合到同一个蛋白质，分子内 FRET 可用来探测蛋白质剪切和构象变化[3]。

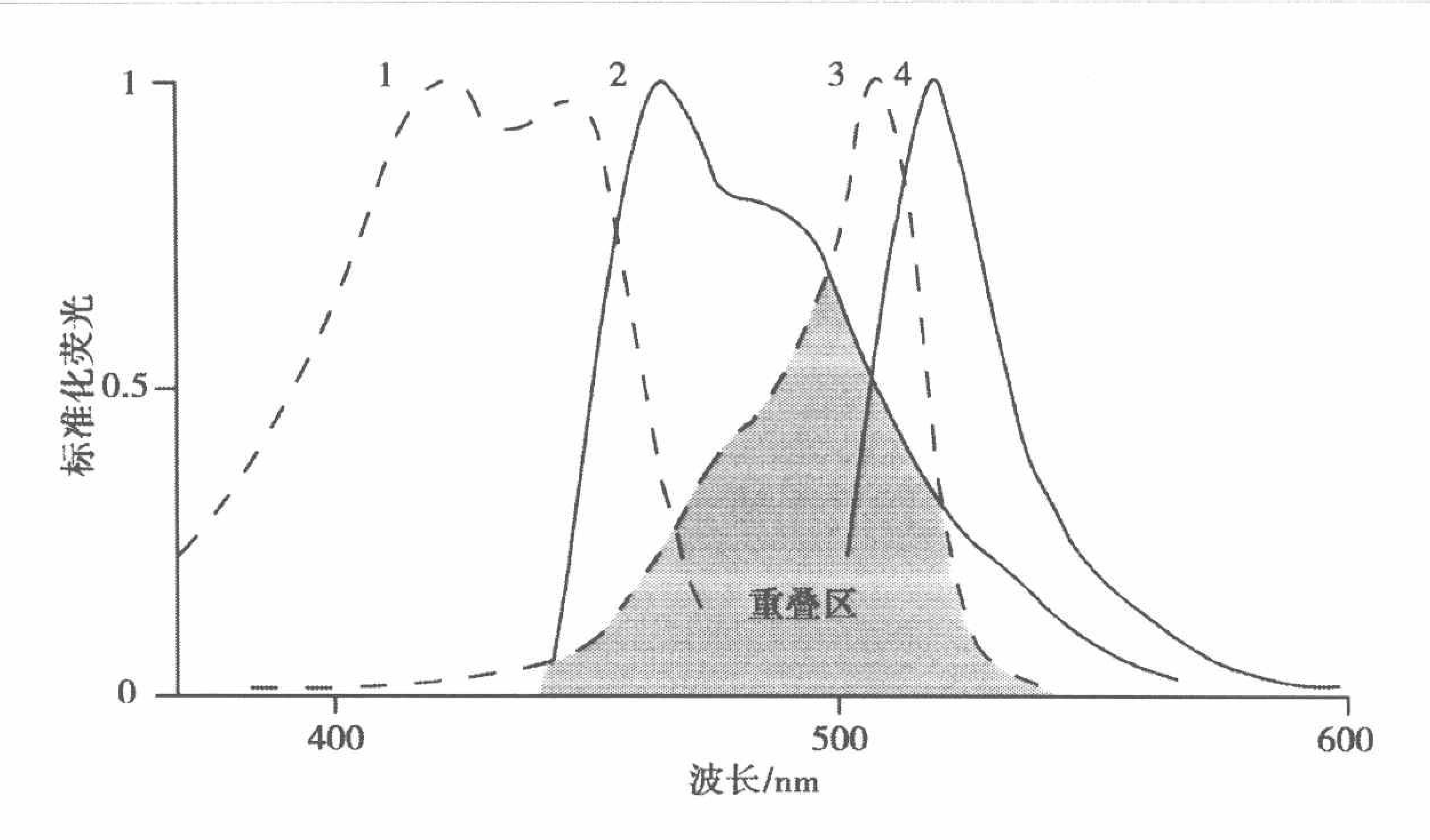


图 4.1 CFP（1 和 2）和 YFP（3 和 4）的发射（实线）与激发（虚线）谱。CFP 的发射谱和 YFP 的激发谱间的重叠使得两个荧光团之间的 FRET 成为可能。

4.1.2 钙变色子的一般设计

使用分子内 FRET 的概念，钙变色子由夹在 CFP 和 YFP 之间的，前后融合的 CaM 的 N 端和 C 端结构域（分别记为 N-CaM 和 C-CaM）以及 CaM 识别序列（CRS）组成[5~7]。CRS 的位置取决于它的结合取向，它或者放在 CaM 的 N 端、CaM 的 C 端，或者放在 N-CaM 和 C-CaM 之间［见图 4.2（A）和注 2］。在 Ca^{2+} 流入之后，CaM 结

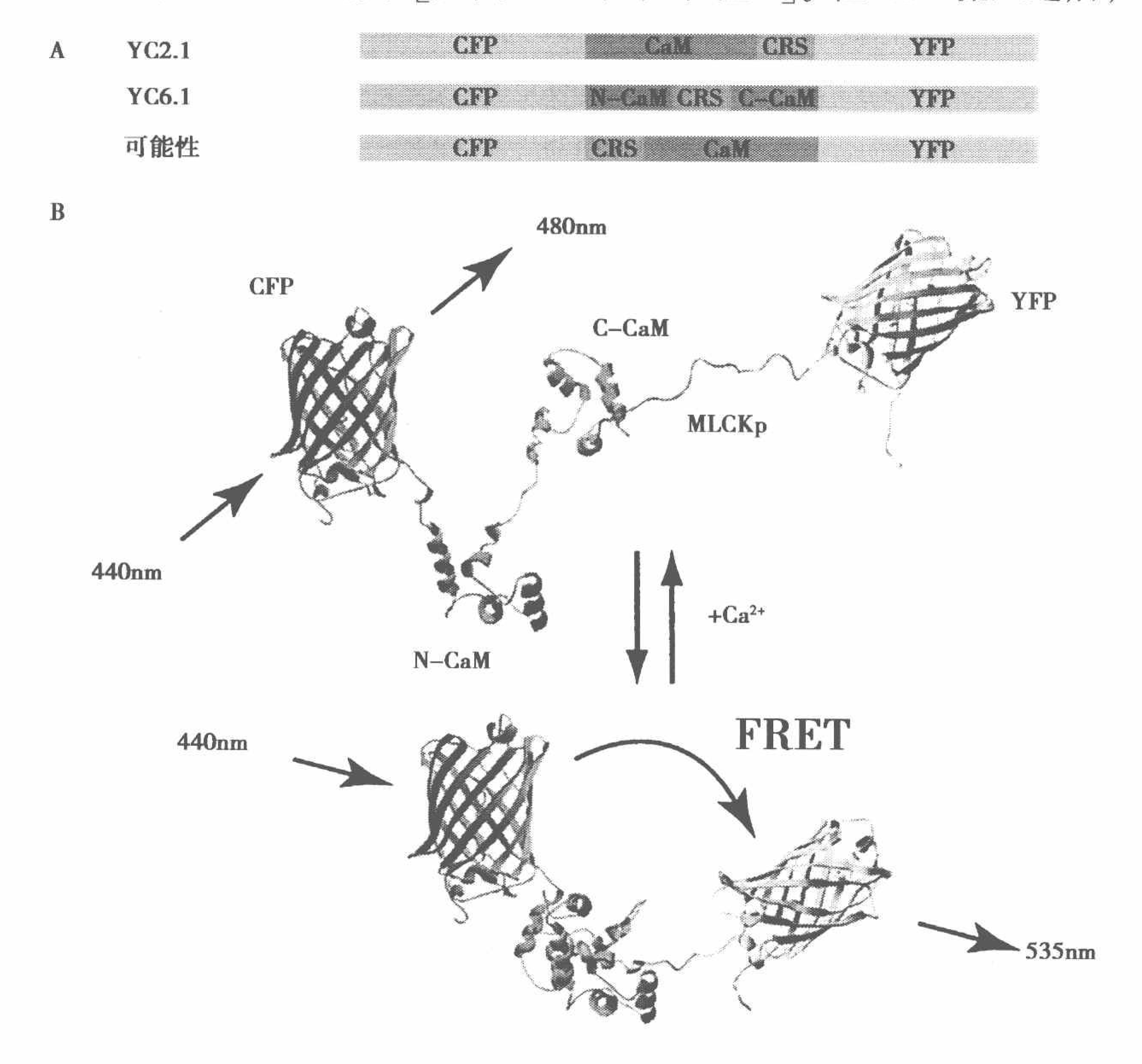


图 4.2 A. 融合 CRS 至 CaM 的几种可能方法。B. YC2.1 可能发生的构象变化的模型。Ca^{2+} 流入之前，两个荧光团相距很远，因而，FRET 很小。Ca^{2+} 流入之后，两个荧光团被 Ca^{2+} CaM-CRS 复合物拉近，使 FRET 增加。

合到 CRS 上，引起构象变化，拉近 CFP 和 YFP，以增加 FRET 效率（见图 4.2B）。已用 CFP-YFP 作 FRET 伙伴，制造出数种黄色的钙变色子（YC），它们具有不同的 Ca^{2+} 亲和力、亚细胞定位和动态范围。本章提供的方法适用于所有已发表的用于 Ca^{2+} 成像的 YC，包括 YC2.1，它基于 CaM 结合到肌球蛋白轻链激酶 CRS 肽上的结构（MLCKp；参考文献[7]）；YC3.1，CaM 的单点突变 E104Q 降低了 Ca^{2+} 亲和力[7]；YC2nu，以核为目标；YC6.1，基于改善了动态范围的，CaM 结合于 CaM 依赖激酶的激酶 CRS 肽（CKKp）的结构[5]。

4.2 材 料

4.2.1 钙变色子的克隆

（1）PRSETB 细菌表达质粒（Invitrogen）：用于 Ni-次氮基三乙酸（NTA）琼脂糖蛋白纯化的 His-tag；用于选择的氨苄青霉素和用于诱导蛋白质表达的异丙基-β-D-硫代半乳糖吡喃糖苷（IPTG）。

（2）pcDNA3 哺乳动物瞬时表达质粒（Invitrogen）：巨细胞病毒启动子和用于细菌选择的氨苄青霉素抗性基因。

4.2.2 生物化学和生物物理鉴定

（1）Luria-Bertani（LB）培养基：10g 胰蛋白胨，5g 酵母提取液和 5g 氯化钠；加入水至终体积 1L。

（2）IPTG。

（3）苯甲基磺酰基氟化物。

（4）Ni-NTA 琼脂糖（Qiagen）。

（5）大肠杆菌菌株 BL21（DE3）。

（6）超声发生器。

（7）EGTA 缓冲液：100mmol/L 氯化钾，50mmol/L pH 7.4 的 HEPES 和 10mmol/L EGTA。

（8）氯化钙缓冲液：100mmol/L 氯化钾，50mmol/L pH 7.4 的 HEPES，10mmol/L EGTA；和 10mmol/L 氯化钙。

（9）荧光光谱仪（Shimadzu）。

4.2.3 钙变色子的成像

（1）含 10%胎牛血清（FBS）的 Dulbecco 的修饰 Eagle 培养液（DMEM）。

（2）含 Ca^{2+} 的 Hank's 平衡盐溶液（Gibco）。

（3）37℃ 二氧化碳培养箱。

（4）HeLa 细胞株。

（5）GeneJuice（Novagen）。

（6）奥林巴斯 IX70 倒置表面荧光显微镜。

（7）奥林巴斯氙灯。

(8) MicroMax 1300YHS 电荷耦合器件（CCD）照相机和 Metafluor 4.5r2 软件控制的 Sutter Lambda 10-2 滤镜转换器。

(9) CFP-YFP FRET 滤镜组（Omega Optical；见注 3）：440AF21 激发滤镜（CFP 激发），455DRLP 二色镜，480AF30 发射滤镜（CFP 发射）和 535AF26 发射滤镜（YFP 发射）。

(10) 中性密度滤镜（ND）组（Omega Optical）。

(11) UApo×40 油 Iris/340 物镜（奥林巴斯）。

(12) U-MNIBA 带通直角镜单元（奥林巴斯）。

(13) 组胺，离子霉素，EGTA 和 BAPTA-AM（Sigma）。

(14) 35mm 直径玻璃底皿（Maltek）。

4.3 方　　法

本文中描述的方法涵盖了变色子在细菌及真核细胞里的克隆与表达（见 4.3.1），变色子的生物化学和生物物理特征鉴定（见 4.3.2），以及通过 FRET 技术将变色子用于体内 Ca^{2+} 成像（见 4.3.3）。

4.3.1 变色子的克隆

使用 5 个不同的部件：CFP、YFP、N-CaM、C-CaM 和 CRS，用标准的重组 DNA 方法[8]模块式组装钙变色子。CFP 和 YFP 质粒可从 Clontech 购得；CaM 来自 Dr. Ikura；CRS 用寡核苷酸链反应合成。为制取钙变色子的 DNA 操控不在这里详细讲述，因为没有适合所有可能性的单一技术（见注 4）。钙变色子制成之后，融合片段克隆到 PRSETB 质粒中，以表达足够蛋白，完成生物化学和生物物理鉴定。我们把融合片段克隆到pcDNA3质粒中（见注 5），以供体内 Ca^{2+} 成像实验用的瞬时哺乳细胞表达。

4.3.2 生物化学和生物物理鉴定

钙变色子在能够用于体内 Ca^{2+} 成像之前，先提纯（见 4.3.2.1 和注 6）并完成体内荧光实验，以确定其动态范围（见 4.3.2.2）和钙离子结合曲线（见 4.3.2.3）。

4.3.2.1 钙变色子从细菌细胞中提纯

(1) 用标准的分子生物学方法以细菌质粒转染 BL21（DE3）细胞[8]。

(2) 把细胞平铺在含氨苄青霉素的 LB 平板上，37°C 保温过夜。

(3) 选择一单菌落，37°C，在 100ml 含 100μmol/L 氨苄青霉素的 LB 培养液中生长。

(4) 在 600nm 光学密度为 0.7 时，用 0.5mmol/L IPTG 诱导 3h。

(5) 以 3000*g* 离心细胞 30min。

(6) 重悬细胞于 10ml，pH 7.4，含 10%甘油、100mmol 氯化钾、1mmol/L 氯化钙和 1mmol/L 苯基甲基磺酰基氟化物的 50mmol/L HEPES 中。

(7) 用 0.5 英寸①超声头以最大功率，10%工作周期，超声 4 次，每次 4min。两次超声之间，溶液降温 5min。

① 1 英寸＝0.0254m。

(8) 以 30 000g 离心细胞碎片 20min。

(9) 在 4°C 轻轻地混合上清液与 1ml Ni-NTA 琼脂糖浆液 30min。

(10) 以 10ml 含 100mmol/L 氯化钾和 5mmol/L 咪唑 pH 7.4 的 50mmol/L HEPES 洗柱。

(11) 用 1ml 含 100mmol/L 氯化钾和 100mmol/L 咪唑的 50mmol/L HEPES 洗脱。不需要切掉 His-tag，因为它并不干扰荧光特性。

(12) 在 4°C，用 2L 含 100mmol/L 氯化钾，pH 7.4 的 50mmol/L HEPES 透析样品。

4.3.2.2 荧光谱分析

FRET 效率的变化（或者，在钙变色子的情形下，Ca^{2+} 浓度的变化）常常可以从发射比的变化观察到（R 为受体发射的峰值强度除以给体的峰值发射强度）。Ca^{2+} 指示子的动态范围定义为最大比率 R_{max} 除以最小比率 R_{min}。较大的动态范围在体内 Ca^{2+} 指示上更有效。

(1) 在装有含 100mmol/L 氯化钾，20μmol/L EGTA 的 50mmol/L HEPES 的 1ml 试管中稀释样品。只要发射信号保持可测，任何稀释因子都可用。在 433nm 激发下，记录从 450～570nm 的荧光发射谱。

(2) 记录空白样品的荧光发射谱。

(3) 从第 (1) 步的记录中减去第 (2) 步的记录，以得到在无 Ca^{2+} 条件下钙变色子的荧光谱。从这个谱中确定 R_{min}。

(4) 在 1mmol/L Ca^{2+} 存在的条件下，重复 (1) ～ (3) 步，以确定 Ca^{2+} 存在时钙变色子的荧光谱。

(5) 图 4.3 显示 YC6.1[5] 在加入 Ca^{2+} 之前和之后的荧光谱，R_{max} 和 R_{min} 分别为 1.1 和 2.4。

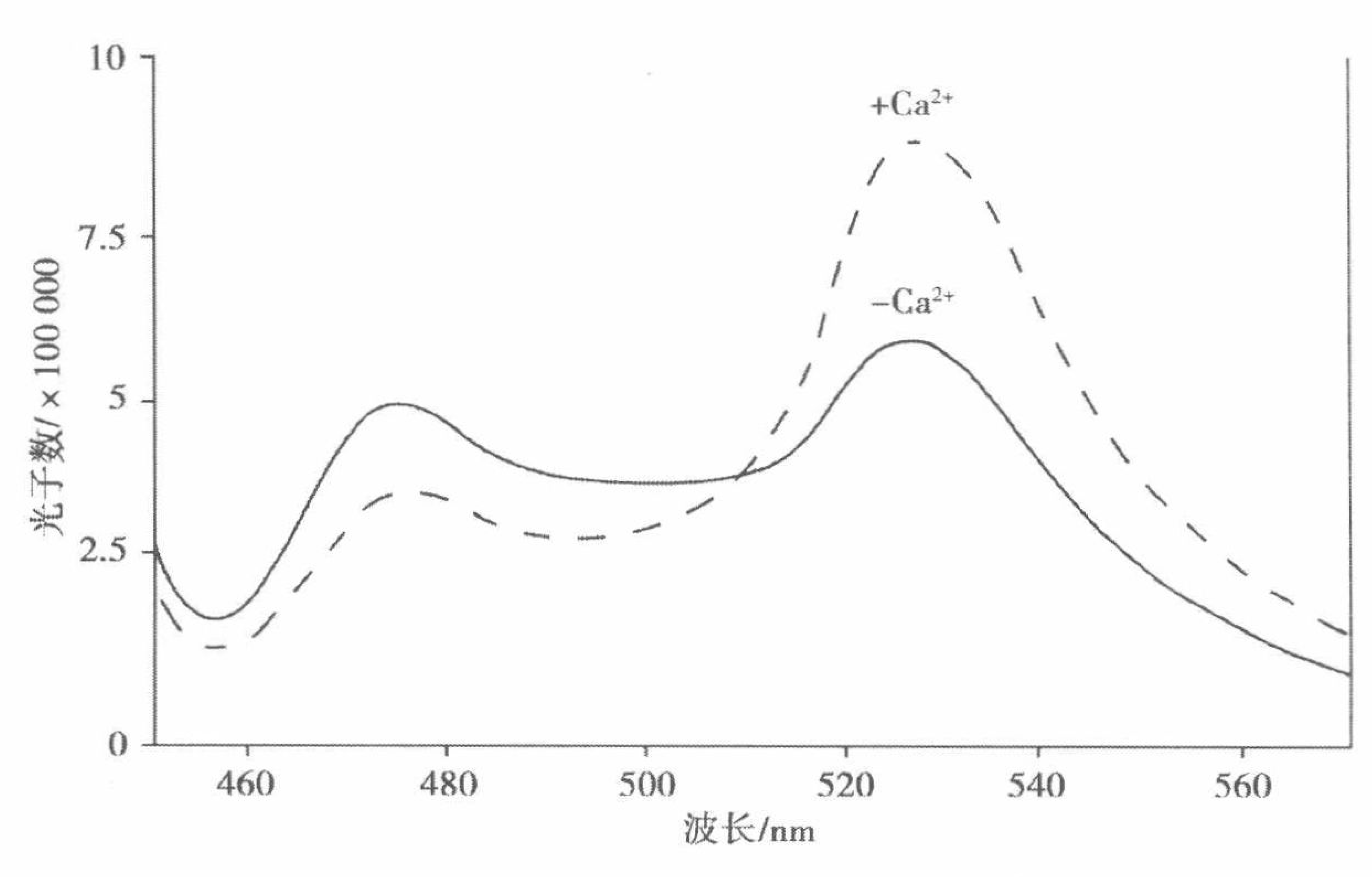


图 4.3 YC6.1 的发射谱。

4.3.2.3　Ca^{2+}结合性质

Ca^{2+}结合曲线用来评估Ca^{2+}指示器能够测量的［Ca^{2+}］的有效范围。Ca^{2+}/EDTA 和Ca^{2+}/EGTA 缓冲液用来作为浓度标准，因为在低［Ca^{2+}］，Ca^{2+}污染会显著地扭曲空白［Ca^{2+}］水平（见注 7）。

（1）在 1ml 试管中，用 EGTA 稀释样品。在 433nm 激发下，记录从 450～570nm 的荧光发射谱。确定激发比。

（2）为得到Ca^{2+}结合曲线，连续地加入氯化钙溶液到样品中，并测定发射比。因为本章描述的实验是在 20℃的 EGTA 和氯化钙缓冲液中完成的，通过解下述二次方程（如在不同条件下，请参考文献［9］）可求得自由钙：

$$([Ca^{2+}]_{free})^2 + \{(10\ 000\ 060.5 - [Ca^{2+}]_{total}) \times [Ca^{2+}]_{free}\} - (60.5 \times [Ca^{2+}]_{total}) = 0$$

（3）为绘出Ca^{2+}结合曲线，将［Ca^{2+}］对发射比作图。图 4.4 显示 YC6.1 的Ca^{2+}结合曲线。

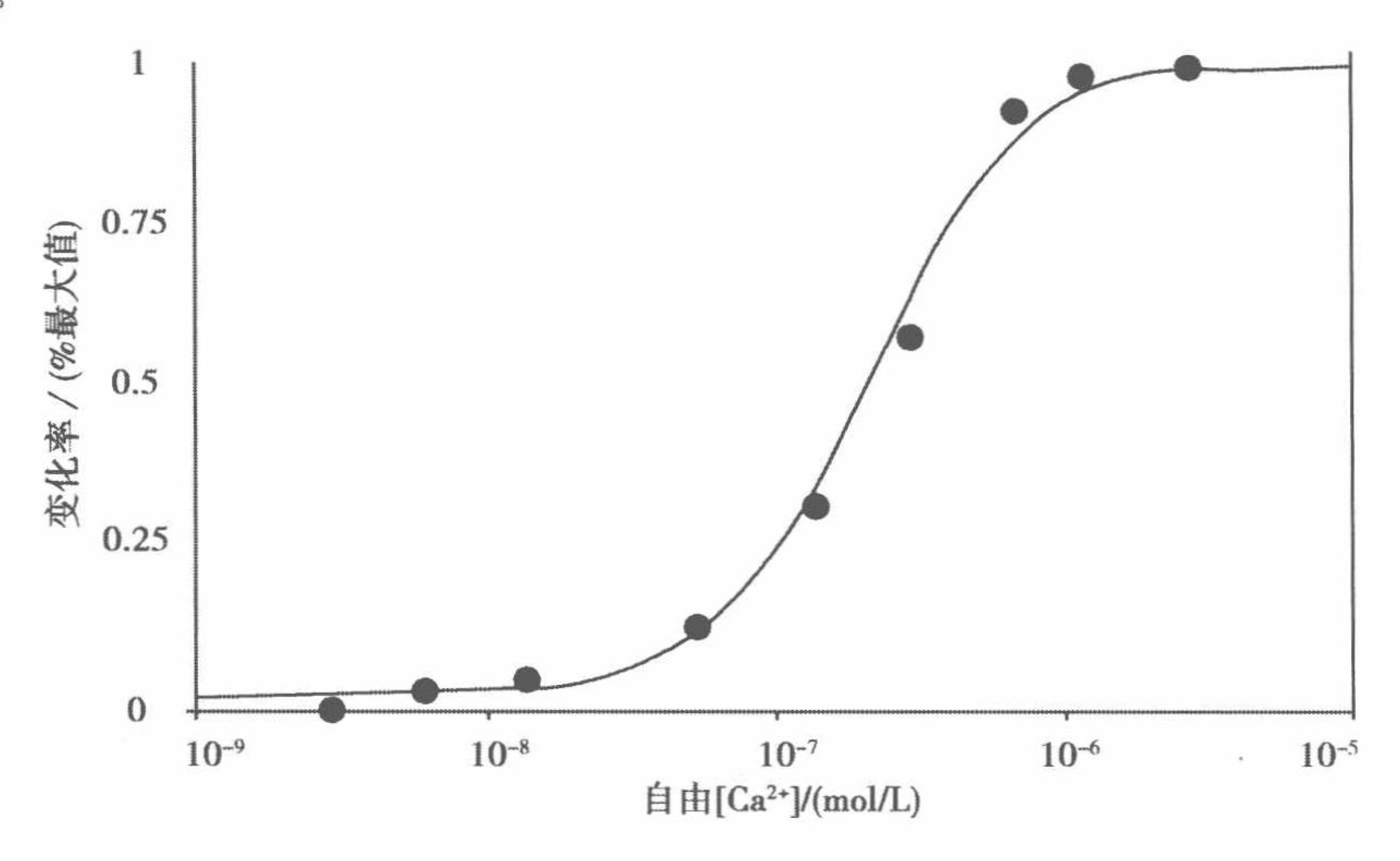


图 4.4　YC6.1 的Ca^{2+}结合曲线实例。

4.3.3　用钙变色子使Ca^{2+}成像

本节描述简单的 HeLa 细胞中Ca^{2+}成像实验；但是只需很小的修正，此方法可应用于其他的黏着细胞和生理条件。分别描述细胞培养液制备（见 4.3.3.1）和数据采集（见 4.3.3.2）。

4.3.3.1　细胞培养液制备

（1）把 HeLa 细胞铺在装有 DMEM-10％FBS 培养液的直径为 35mm 的玻璃皿上。

（2）在 37℃（5％二氧化碳）保育，直到 50％～80％细胞汇合。

（3）使用 GeneJuice（Novagen），将含钙变色子的哺乳细胞表达质粒转染细胞。

（4）6h 以后移除转染混合液，代之以 1.5ml 的 DMEM-10％FBS 培养液。

（5）在 37℃（5％二氧化碳）保育 24h。能完成Ca^{2+}成像实验的细胞已备好。

4.3.3.2 数据采集

实验应该在暗室中完成，以减少背景光。图 4.5 显示我们使用的显微设备。

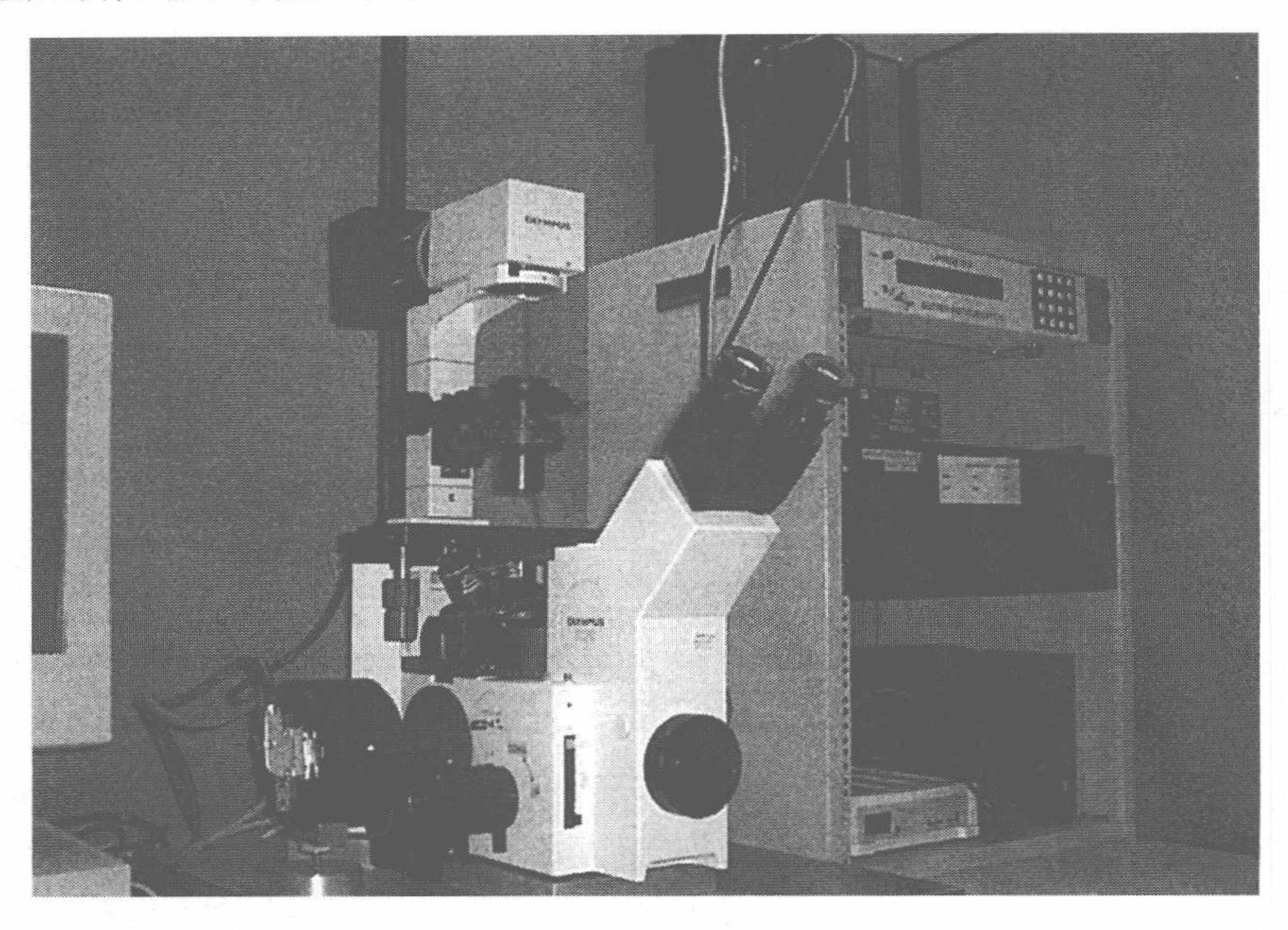


图 4.5 数字荧光显微镜设备。

（1）打开氙灯、显微镜、滤光转换器和计算机。

（2）从培养皿中移除培养基。

（3）用 1ml HBSS（+氯化钙）清洗。

（4）加入 1ml 新鲜的 HBSS（+氯化钙）。

（5）把细胞放在显微镜台面上。

（6）启动 MetaFluor 软件。在采集数据时，MetaFluor 软件控制快门、滤光转换器和照相机。

（7）用目镜挑选 YFP 荧光，以发现表达钙变色子的细胞。把滤光指示转到 U-WNIBA带通直角镜位置，并根据荧光发射水平，使 ND 滤镜在 0.1%～10%。U-WNIBA直角镜对使用目镜快速发现 YFP 荧光细胞很实用，而 10%ND 滤镜是用来防止光褪色。

（8）使用 40×油物镜，把显微视场中心移到外观健康、并呈现强胞质体荧光的细胞上。图 4.6 是显示健康 HeLa 细胞典型荧光分布的画面。

（9）用 MetaFluor 在计算机荧屏上逐幅采集图像，同时调节聚焦直至得到最清晰的图像。

（10）把滤光指示转到 CFP-YFP FRET 直角镜位置。

（11）为监控作为 CFP 对时间的峰值激发的结果的 CFP 和 YFP 的峰值激发，设置如下数据采集条件：时间间隔大于 10s；对 CFP 和 YFP，曝光时间 200ms。MetaFluor 会显示发射比随时间的变化。

（12）在 CCD 图像上选择一块感兴趣的区域。通常这个区域要涵盖感兴趣的细胞。

然后开始数据采集。在感兴趣的区域内的 CFP 和 YFP 的荧光发射强度，和它们的发射比一起，会得到实时监控（见注 8）。

（13）当发射比达到稳衡态，向培养皿中加入 50μl 2mmol/L 组胺，使终浓度大约为 100μmol/L。操作过程要小心，不要移动培养皿。组胺结合到质膜上的受体上，启动信号级联，使 Ca^{2+} 从内质网通过肌糖-1,4,5-三磷酸受体释放[10,11]。这将会引起钙变色子的构象变化，因 YFP 发射强度增加 CFP 强度减弱而被观察到。因而，发射比应该增加。当组胺作用消退后，发射比应该减退到稳衡态水平。图 4.6 显示 MetaFluor 软件在采集数据时的一幅荧屏照片。

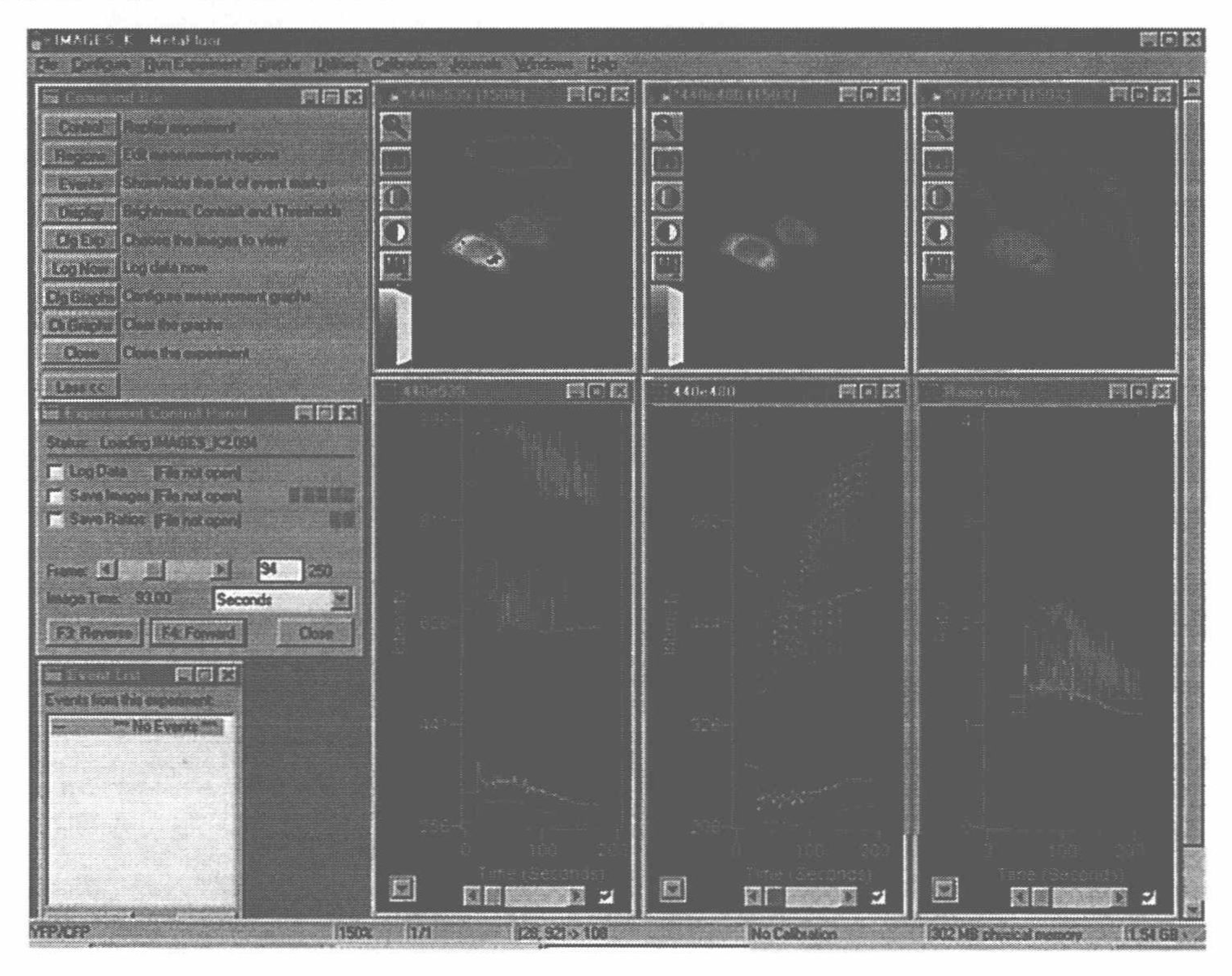


图 4.6　MetFluor 软件的荧屏照像。在 440e535 小图中，标出了观察区。440e535 小图和下面的曲线图画出了 CFP 激发的 YFP 荧光的变化。440e480 小图和下面的曲线画出了 CFP 激发的 CFP 荧光的变化。YFP/CFP 小图和下面的曲线是 440e535 和 440e480 小图和曲线之比。在这个实验中，图形显示 Ca^{2+} 受组胺刺激而（另见彩图）振荡。

（14）为使发射比与胞质体［Ca^{2+}］相关联，必须测定 R_{min} 和 R_{max}，以便把发射比映射到 Ca^{2+} 结合曲线上。加入 50μl 的 20μmol/L 离子霉素，使终浓度近似为 1μmol/L。离子霉素在质膜上把通道打开，以让 Ca^{2+} 渗过。由于培养基的氯化钙饱和，发射比会升高到 R_{max}。为测定 R_{min}，加入 50μl 100mmol/L 的 EGTA 和 600μmol/L BAPTA-AM，使终浓度分别近似为 5mmol/L 和 30μmol/L。发射比应该会降到 R_{min}。

（15）停止数据采集。

4.4 注

1. 钙变色子不是没有它们自身的局限。通常，它们比提到的其他方法有较低的信噪比。但是，正如文中所述，信噪比可通过仔细设计新的建构[5]或通过优化实验条件来改善。

2. 插入 CRS 的最好的地方，依赖于三维结构中，两个 CaM 结构域相对于 CRS 的排布。如果在 CaM-CRS 复合物中，CRS 的 N 端在 CaM 的 C 端的 20Å 之内，CaM-CRS 融合就是有效的[6]。但是，如果 N 端和 C 端都在 CaM 铰链的 10Å 之内，（N-CaM）-CRS-（C-CaM）融合会是最有效的[5]。

3. 如果在自己的应用中只对使用 CFP 和 YFP 感兴趣，CFP-YFP FRET 滤光组合会满足你的需求。如果你想在 FRET 中使用 Discosoma 红色荧光蛋白（dsRed；参考文献 [12] 和 [13]），需要下列来自 Omega Optical 的或类似的双色镜和滤光镜：双色镜 450-520-590TBDR 用于二色镜；440DF21 用于 CFP 激发；510DF23 用于 YFP 激发；575DF26 用于 dsRed 激发。480AF30 用于 CFP 发射；535AF26 用于 YFP 发射；600ALP 用于 dsRed 发射。这一滤镜组允许你激发和接收单独来自 CFP、YFP 和 dsRed 的光，但效率会有降低。

4. 在钙变色子的建构中，我们切除了 CFP 的 C 端 11 个氨基酸（形成 GFP 极小域），以减少荧光团的相对翻滚。甚至，把氨基乙酸-甘氨酸连接引入 CaM 和 CRS 融合蛋白中，以增加 Ca^{2+} CaM-CRS 复合物的构象自由度。

5. 在为插入 PRSETB 或 pcDNA3 而用聚合酶链反应扩增钙变色子 DNA 时，记住钙变色子的 5′和 3′引物会有内引物位点，因为，CFP 和 YFP 只有几个点突变的差别。需从琼脂糖胶上纯化大小正确的 DNA 片段。

6. 如果你的实验中钙变色子降解或在细菌细胞中不溶，可以用哺乳细胞降解物完成鉴定。为收获产物，细胞应生长在直径为 10cm 的培养皿中。转染 24h 后，把细胞刮出培养皿并在低渗的破解缓冲液（50mmol/L HEPES，pH 7.4 和 100mmol/L 氯化钾；5mmol/L 氯化镁和 0.5% Triton X-100）中破碎。然后离心混合物以去除碎片。下一步，在 2L 缓冲液（50mmol/L HEPES，pH 7.4 和 100mmol/L 氯化钾）中透析上清液。最后，样品可用于鉴定，如 4.3.2.2 和 4.3.2.3 节所述。

7. Ca^{2+} 结合曲线的精度取决于 Ca^{2+}/EGTA 缓冲液的精确配制[9]。

8. 由于光漂白作用，实验过程中发射强度会增加；但是，如果 FRET 不变的话，发射强度应保持为常数。为减少光漂白，减少曝光时间和激发光强。以“Binning”方式读取 CCD 照片数据时，会将许多点的数据相加，所以在维持高信噪比的同时，只需较少的光。

致谢 K. T. 感谢加拿大健康研究所和 Ontario Student Opportunity Trust Funds 奖学金。本工作得到癌症研究会的基金支持。M. I. 是加拿大健康研究所研究员。

参考文献

1. Takahashi, A., Camacho, P., Lechleiter, J. D., and Herman, B. (1999) Measurement of intracellular Ca^{2+}. *Physiol. Rev.* 79, 1089-1125.

2. Miyawaki, A. and Tsien, R. Y. (2000) Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. *Methods Enzymol.* 327, 472-500.

3. Truong, K. and Ikura, M. (2001) The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes *in vivo*. *Curr. Opin. Struct. Biol.* 11, 573-578.

4. Tsien, R. Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem.* 67, 509-544.

5. Truong, K., Sawano, A., Mizuno, H., et al. (2001) FRET-based *in vivo* Ca^{2+} imaging by a new calmodulin-GFP fusion molecule. *Nat. Struct. Biol.* 8, 1069-1073.

6. Miyawaki, A., Llopis, J., Heim, R., et al. (1997) Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* 388, 882-887.

7. Miyawaki, A., Griesbeck, O., Heim, R., and Tsien, R. Y. (1999) Dynamic and quantitative Ca^{2+} measurements using improved chameleons. *Proc. Natl. Acad. Sci. USA* 96, 2135-2140.

8. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual.* 2nd ed., Cold Spring Harbor Laboratory Press, New York.

9. Tsien, R. and Pozzan, T. (1989) Measurement of cytosolic free Ca^{2+} with quin2. *Methods Enzymol.* 172, 230-262.

10. Bootman, M. D., Cheek, T. R., Moreton, R. B., Bennett, D. L., and Berridge, M. J. (1994) Smoothly graded Ca^{2+} release from inositol 1, 4, 5-trisphosphate-sensitive Ca^{2+} stores. *J. Biol. Chem.* 269, 24783-24791.

11. Zamani, M. R. and Bristow, D. R. (1996) The histamine H1 receptor in GT1-7 neuronal cells is regulated by Ca^{2+} influx and KN-62, a putative inhibitor of Ca^{2+}/calmodulin protein kinase II. *Br. J. Pharmacol.* 118, 1119-1126.

12. Mizuno, H., Sawano, A., Eli, P., Hama, H., and Miyawaki, A. (2001) Red fluorescent protein from Discosoma as a fusion tag and a partner for fluorescence resonance energy transfer. *Biochemistry* 40, 2502-2510.

13. Matz, M. V., Fradkov, A. F., Labas, Y. A., et al. (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* 17, 969-973.

5　人造锌指蛋白的设计和合成

Wataru Nomura and Yukio Sugiura

概要　在 DNA 结合模体中，$(Cys)_2$ $(His)_2$ 类型的锌指模体具有大的操控潜力。为新的 DNA 结合蛋白设计，锌指模体提供了有吸引力的框架。特别是为产生新的、具有全新的 DNA 结合特性的，如长 DNA 链识别、DNA 弯折和 AT 富集序列识别——人造锌指蛋白，基于结构的设计技巧是特别有价值的。在此，基于最新的实验结果，描述为设计多重锌指蛋白以用于目标 DNA 序列识别、DNA 弯折锌指蛋白、$(His)_4$ 类型的锌指蛋白和 AT 识别锌指蛋白的新技巧。

关键词　锌指；多指；DNA 弯折；GC 识别；AT 识别；$(Cys)_2$ $(His)_2$ 型；$(His)_4$ 型；人工蛋白；螺旋取代；蛋白质设计

5.1　介　　绍

人造锌指蛋白在生物医学研究和基因疗法中有重要应用，而且是生物化学和分子生物学研究的新工具[1~3]。特别是 $(Cys)_2$ $(His)_2$ 型锌指模体（真核细胞中最常见的 DNA 结合模体）为新的 DNA 结合蛋白设计提供通用的、引人瞩目的框架。对某些 $(Cys)_2$ $(His)_2$ 型锌指蛋白-DNA 复合物揭示出这一模体与 DNA 结合的下列特征：

(1) 锌指结构用一特异的连接子反复连接；

(2) 每一个锌指结构结合到 DNA 的一个三碱基对位置上，以其螺旋对着主沟槽；

(3) α 螺旋的每一锌指单元的 4 个关键位置上的氨基酸残基与 DNA 碱基在特定位置上 1∶1 的接触；

(4)（锌指）多肽的整体排列是与 DNA 主要相互作用股反向平行的[4~7]。

基于这些特征，科学家们成功地制造了某些锌指蛋白，如长序列或 AT 序列结合锌指。这样设计的 DNA 结合蛋白，可以期待拥有具高亲和力和特异性的唯一结合序列。还设计了新的锌指蛋白，如 DNA 弯折锌指和 $(His)_4$ 型锌指。这些人工的锌指蛋白，在不久的将来，为基因工程和基因组特异的转录开关提供了巨大的希望和有用的工具。

本章内容集中在设计和制备新的人工锌指蛋白，也就是九锌指蛋白（见 5.3.2）、DNA 弯折锌指蛋白（见 5.3.3）、$(His)_4$ 型锌指蛋白（见 5.3.4）和 AT 识别锌指蛋白（见 5.3.5）。

5.2　材　　料

(1) pET3b 表达系统（Novagen）。

(2) 互补 DNA 编码 Sp1 锌指。

(3) 大肠杆菌菌株 DH5α 和 BL21 (DE3) pLysS。

(4) 正文中描述的寡核苷酸引物。

(5) 限制酶，T7 DNA 聚合酶和 T4 DNA 连接酶。

(6) DNA 测序仪 (Amersham Biosciences)。

(7) Luria-Bertani 培养基：10g 胰蛋白胨、5g 酵母抽提物和 10g 氯化钠，加水至终体积 1L。

(8) 氨苄青霉素。

(9) 异丙基-β-D-硫代-吡喃半乳糖苷 (IPTG)。

(10) 超声裂解设备。

(11) 磷酸缓冲生理盐水 (PBS)：10mmol/L 磷酸缓冲液，pH 7.6，130mmol/L 氯化钠和 2.7mmol/L 氯化钾。

(12) 十二烷基硫酸钠聚丙烯酰胺凝胶电泳 (SDS-PAGE) 设备。

(13) 色谱设备。

(14) 高 S 色谱柱 (Bio-Rad)。

(15) UNO-S1 色谱柱 (Bio-Rad)。

(16) Superdex 75 (Amershan Biosciences)。

(17) 50mmol/L Tris-氯化氢，pH 8.0，50mmol/L 氯化钠和 1mmol/L 二硫苏糖醇。

(18) PAGE 设备。

(19) DNA 测序设备。

(20) 9-芴甲氧羰基固相合成设备。

(21) 三氟乙酸：乙醇 (95：5) 溶液。

(22) 高效液相色谱设备。

(23) μBonedesphere$_5$ C_4-300 (19mm×150mm) 柱。

(24) 飞行时间质谱设备。

(25) 圆二色 (CD) 谱设备。

(26) 紫外-可见 (UV-VIS) 光谱仪。

(27) 核磁共振 (NMR) 设备。

5.3 方　　法

本节描述的方法大概包括：

(1) 设计的锌指蛋白的表达和纯化。

(2) 人工锌指设计技巧。

(3) 人工锌指构建步骤。

(4) 锌指结构和金属配位的确认。

5.3.1 人工锌指蛋白的表达和纯化

设计的人工锌指蛋白的表达和纯化方法在 5.3.1.1 和 5.3.1.6 节描述，包括：

(1) 表达载体 pEV-3b 的构建；

（2）蛋白质以可溶和不可溶形式表达；

（3）用阳离子交换和凝胶过滤色谱纯化蛋白；

（4）失活结构的再折叠；

（5）锌指结构和金属配位的确认。

5.3.1.1 pEV-3b 表达载体

pEV-3b 表达载体是基于 pET3b 表达载体（Novagen；图 5.1）。这个载体被改造过，以生成有用的多克隆位点，用于研究锌指基因。

（1）从 pET3b 中切除 *Eco*RI/ *Hind*III 降解片段。

（2）为了切除 *Eco*RI 和 *Hind*III 位点，将退火的寡核苷酸 5′-AATTGTCATGTTTGAC-3′和 5′-AGCTGTCAAACATGAC-3′插入 *Eco*RI/ *Hind*III 降解的 pET3b。产生的质粒记为 pET3b′。

（3）制备包括限制性酶 *Afl*II、*Bam*HI、*Eco*RI、*Hind*III 和 *Sma*I 酶切位点的双链寡核苷酸。这两条链的序列为 5′-TATGGATCCCGGGAATTCAAGCTTAAGC-3′和 5′-TCAGCTTAAGCTTGAATTCCCGGGATCCA-3′。

（4）用 *Nde*I 和 *Bpu*1102I 降解，并将这一片段插入 pET3b′，得到质粒 pEV-3b。

得到的质粒 pEV-3b 用来表达设计的锌指蛋白。把锌指基因作为 *Bam*HI/ *Eco*RI 片段插入类似地降解的 pEV-3b，只用一步就构建了人工锌指的表达载体。

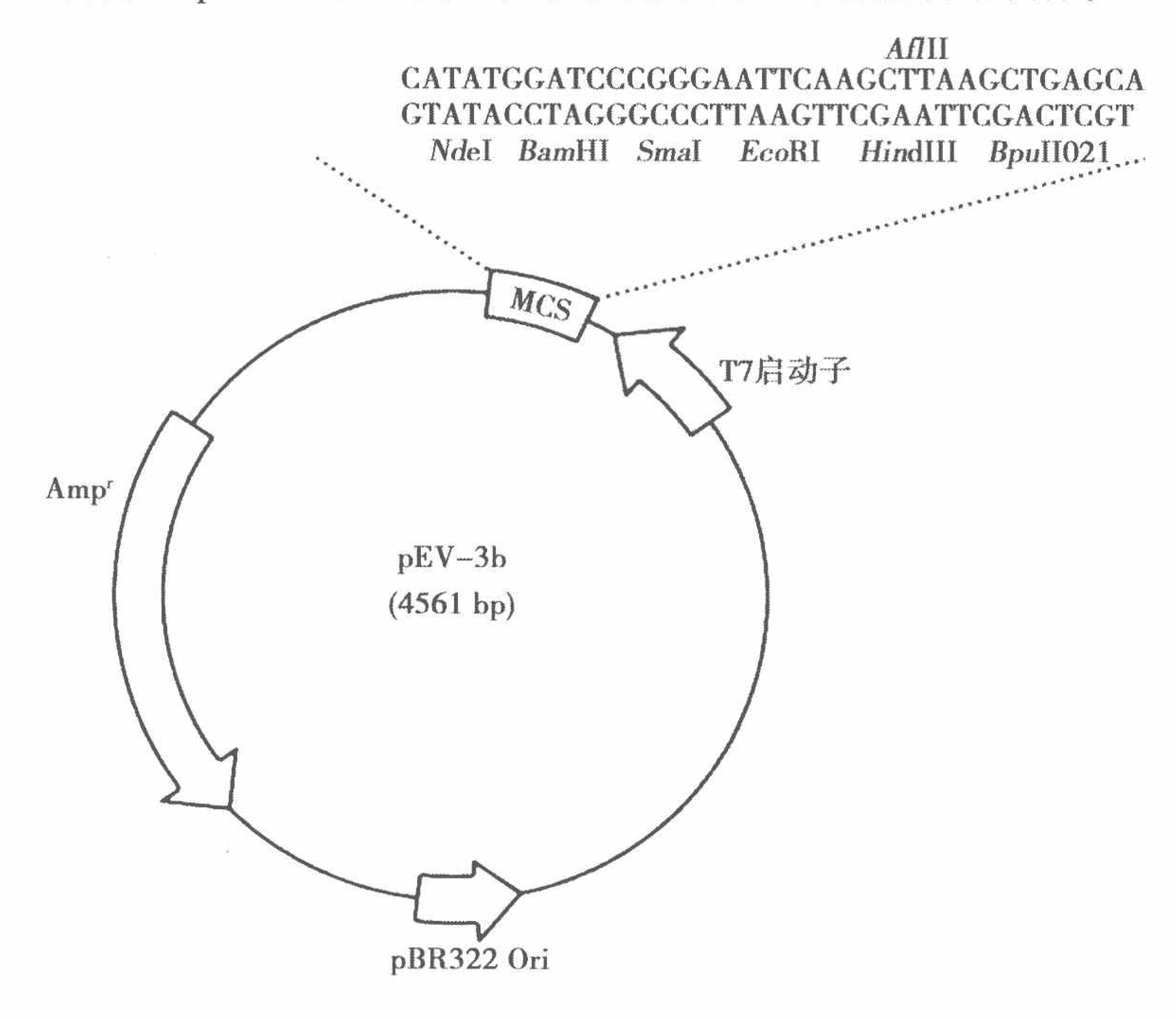


图 5.1 pEV-3b 表达载体的框架图。画出了多克隆位点的序列，限制性酶切位点也在序列中标出。

5.3.1.2 设计的人工锌指蛋白的表达

（1）把带有设计的锌指蛋白基因的 pEV-3b 质粒转化到大肠杆菌菌株 BL21（DE3）

pLysS 中。

(2) 在 37°C 于 Luria-Bertani 培养基中培养细胞。

(3) 当在 600nm 的光密度为 0.6 时加入 1mmol/L IPTG。

(4) 培养液在 20°C 保育 8～12h。为了表达溶解状态的蛋白质，这个温度很重要。

5.3.1.3 设计的人工锌指蛋白的纯化

纯化步骤在 4°C 下完成。

(1) 收获大肠杆菌细胞，重悬，并在 PBS 中裂解。

(2) 离心后，用阳离子交换色谱（高 S 和 UNO S-1；Bio-Rad）和用 Tris-氯化氢缓冲液凝胶过滤（Superdex 75；Amersham Biosciences），纯化溶有锌指蛋白的上清液。

(3) 用 SDS-PAGE 确认蛋白纯度。

为进行金属取代实验，不溶形式的锌指蛋白也在离心后从大肠杆菌细胞小团中纯化出来：

(1) 在含 8mmol/L 尿素和 10mmol/L 螯合剂（EDTA 或 1,10-邻二氮杂菲）的 PBS 中裂解细胞小团。

(2) 根据 5.3.1.3 小节第 (2) 步的描述，以同样步骤纯化。

(3) 在 65°C 加热纯化蛋白 30min，并在含 125μmol/L 氯化锌、硝酸镍、氯化镉、硝酸钴或硫酸铜的 10mmol/L Tris 缓冲液中逐步降温，使之重新折叠。

5.3.1.4 CD 测量

锌指蛋白的 CD 谱用 Jasco J-720 分光偏振计记录。20℃，使用 pH 8.0 的含有 50mmol/L 氯化钠的 Tris-氯化氢缓冲液。容器须有盖，1mm 光程，氮气环境。所有的谱平均扫描 8～16 次。使用 Jasco 软件校准谱的基线并降低噪声。

5.3.1.5 UV-VIS 吸收谱

UV-VIS 吸收谱用 Beckman Coulter DU7400 二极管阵列光谱仪，20℃，使用 pH 7.5 的含 50mmol/L 氯化钠的 10mmol/L Tris-氯化氢缓冲液。容器须有盖，1cm 光程。Co (II) -取代的锌指复合物用氯化钴滴定得到。多肽在任意条件下用 Co (II) 饱和。所有的谱都用 $\varepsilon=A/(l\cdot c)$，式中 ε 是消光系数 {/ [(mol/L) /cm]}，l 是容器的光程 (cm)，c 是肽浓度 (mol/L)。

5.3.1.6 NMR 实验

在 1.5mol Zn (II) 离子存在的条件下，单指结构域和 Zn (II) 的复合物在 90% H_2O/10%D_2O 和 D_2O (25mmol/L Tris-d_{11}，pH 5.7) 中，以 5mmol/L 浓度制备。所有的 NMR 谱都在 JOEL Lambda-600 谱仪上记录。

(1) 核 Overhauser 增强谱（NOESY）数据采集，先用选择性水预饱和（交替的延迟与进动以裁剪激发脉冲；一种水峰信号抑制技术，缩写为 DANTE；原文 mutation 当为 nutation 之误——译者注），随后温度定在 303K，各使用混合时间分别为 100ms、200ms 和 300ms 的标准 NOESY 脉冲序列。

(2) 采集总相关谱，用 80ms MLEV-17 自旋锁闭时长，在 303K 用梯度修整脉冲以压制水信号。

谱的典型采集条件是每个 t1 值（原文 valve 当为 value 之误——译者注）扫描 24 次，共 1024 个 t1 值，并且在直接维度收集 2048 个复数点。在两个维度的自由感应衰减被乘以相移正弦钟形限形函数、零填充，并经傅里叶变换为 2048×2048 矩阵。序列共振指认用标准的总相关谱和 NOESY 过程确定[8]。

5.3.2 6和9锌指蛋白

本节描述为识别长 DNA 序列所需的多锌指蛋白的设计策略和构建。

5.3.2.1 蛋白质设计策略

全新的 6 和 9 锌指蛋白（Sp1ZF6 和 Sp1ZF9）是从转录因子 Sp1 的 3 锌指模体创造出来的（图 5.2）。通过用 Krüppel 型连接子连接 2 或 3 个 Sp1 锌指结构域，构建出了这些蛋白质（见注 1 和注 2）。

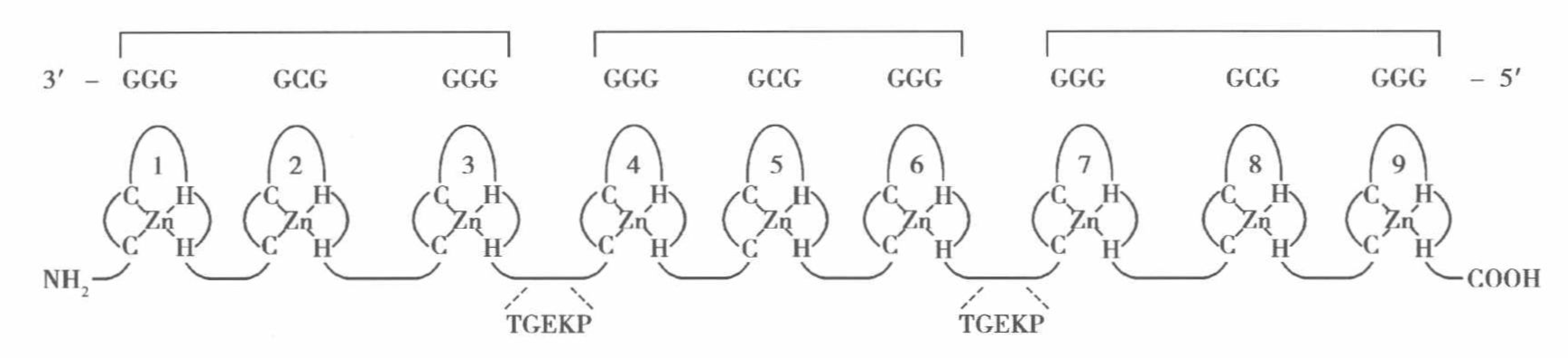


图 5.2 图示人造 9 锌指蛋白，Sp1ZF9。DNA 与蛋白质的结合位点标示在 9 锌指图示之上（见注 2）。Krüppel 型连接序列及其他氨基酸用单字母码表示。

5.3.2.2 *Sp1ZF6* 和 *Sp1ZF9* 基因的构建

编码 Sp1 的 3 锌指区域的基因［pUC-Sp1（530～623）］按如先前描述过的步骤构建[9]。

(1) 合成编码 Krüppel 型连接子（TGEKP）的一段寡核苷酸（84bp）为 *Bam*HI/*Sty*I 片段，并将其插入 pUC-Sp1（530～623）。

(2) 切出 *Eco*47III 片段（264bp），并将其插入简单降解过的 pUC-Sp1（530～623）中。改变后的质粒重新命名为 pUC-Sp1ZF6。

(3) 用 *Age*I 位引物对，即用 5′-ACCGGTGAAAAACCGCATATTTGCCACATAC-3′为编码链，并用 5′-CGGTTTTTCACCGGTGTGGGTCTTGATATG-3′为非编码链，将两个 *Age*I 位分别连接到编码 Sp1ZF9 的中间 *Sp1* 基因的 5′端和 3′端。

(4) 把得到的用 *Age*I 修饰过的片段连接到 pUC-Sp1ZF6 的 *Age*I 位置。连接 2 或 3 个 Sp1 片段间的 *Age*I 酶切位点编码氨基酸 TG，它是连接肽 TGEKP 的一部分。

(5) 用 DNA 测序确认所有序列。

(6) 作为 *Bam*HI/*Sty*I 片段切出 Sp1ZF6 和 Sp1ZF9 的 DNA 片段，并将其插入简单降解过的质粒 pEV-3b 中（见 5.3.1.1；参考文献［10］）。

5.3.3 DNA 弯折锌指蛋白

5.3.3.1 和 5.3.3.2 节描述 DNA 弯折锌指的设计和构建。

5.3.3.1 蛋白设计步骤

通过连接 Sp1 的 2 个 3 锌指，构建出了新的 6 锌指蛋白，包括多聚甘氨酸连接子、Sp1ZF6 $(Gly)_4$、Sp1ZF6 $(Gly)_7$（图 5.3）和 Sp1ZF6 $(Gly)_{10}$。交换 Sp1ZF6 的 TGEKP 和 $(Gly)_n$（n=4、7 或 10）连接子，可以构建 Sp1ZF6 $(Gly)_n$（见注 3 和注 4）。

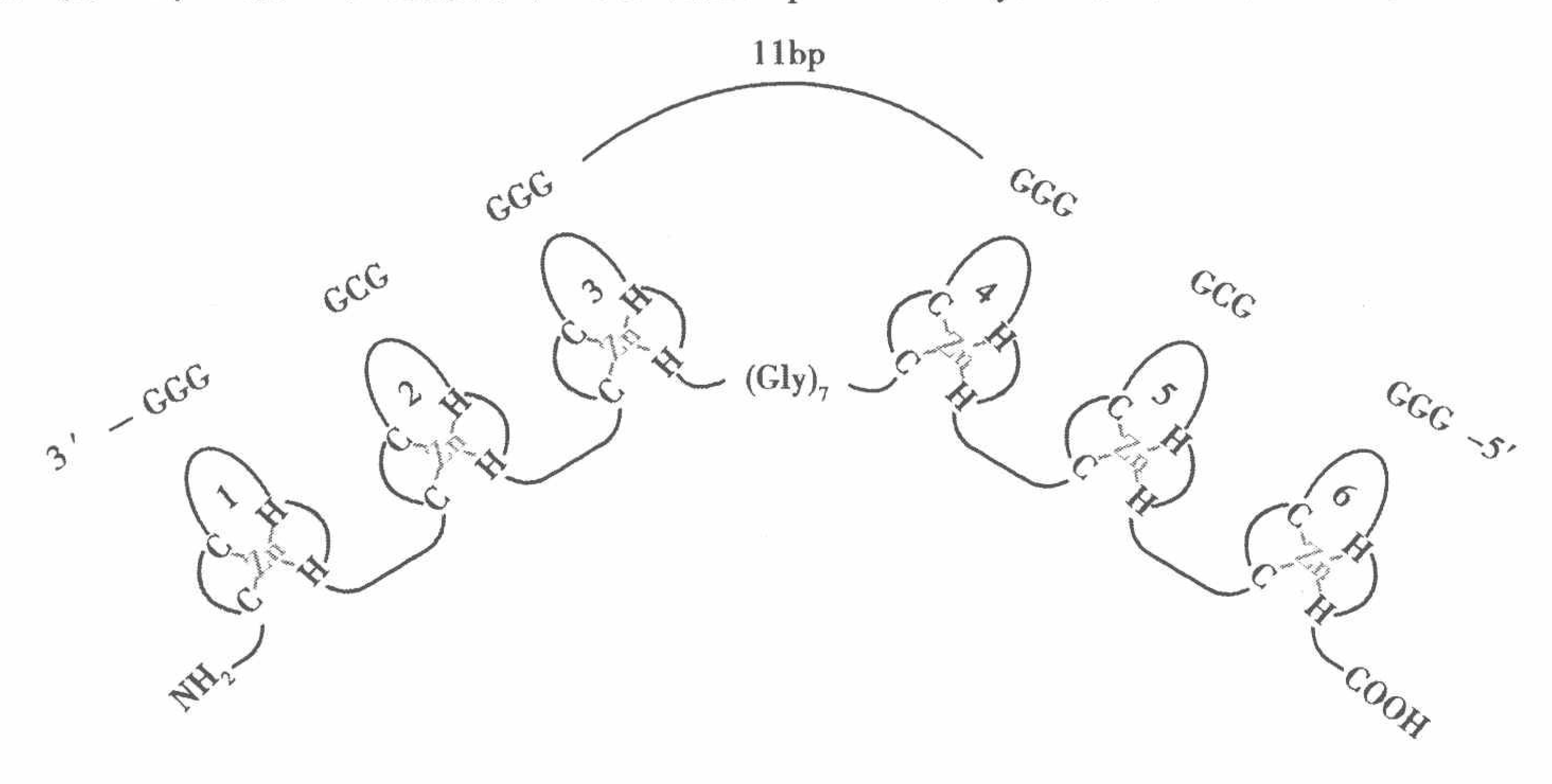


图 5.3 Sp1ZF6 $(Gly)_7$ 的框架图。被 6 锌指蛋白所弯折的 DNA 序列显示在图的上部，氨基酸残基用单字母符号标明。

5.3.3.2 Sp1ZF6 $(Gly)_n$ 的构建

包含有连接序列的合成寡聚核苷酸 AflII/MfeI 是从 Amer sham Biosciences 购置的。经过退火，这些寡聚核苷酸片段被插入到 pUC-Sp1ZF6 载体中。这样得到的质粒被重新命名为 Sp1ZF6 $(Gly)_n$（n=4，7 或 10）。这些编码蛋白质的 DNA 片段在 *Bam* HI 和 *Eco* RI 位点被剪切下来，然后插入一个被同样酶切的质粒，pEV-3b（见 5.3.1.1；参考文献 [11]）

5.3.4 $(His)_4$ 型锌指蛋白

5.3.4.1 和 5.3.4.2 节描述 $(His)_4$ 型锌指蛋白的设计技巧、多肽合成和基因构建。

5.3.4.1 蛋白质设计策略

通过对 3 $(Cys)_2$ $(His)_2$ 型锌指蛋白 Sp1 做突变 CysHis，创造出了一个新的 $(His)_4$ 型锌指蛋白（图 5.4）。实质上，存在 $(Cys)_2$ $(His)_2$—、$(Cys)_3$ (His) —、$(Cys)_4$—和 $(Cys)_6$—锌指蛋白（见注 5）。

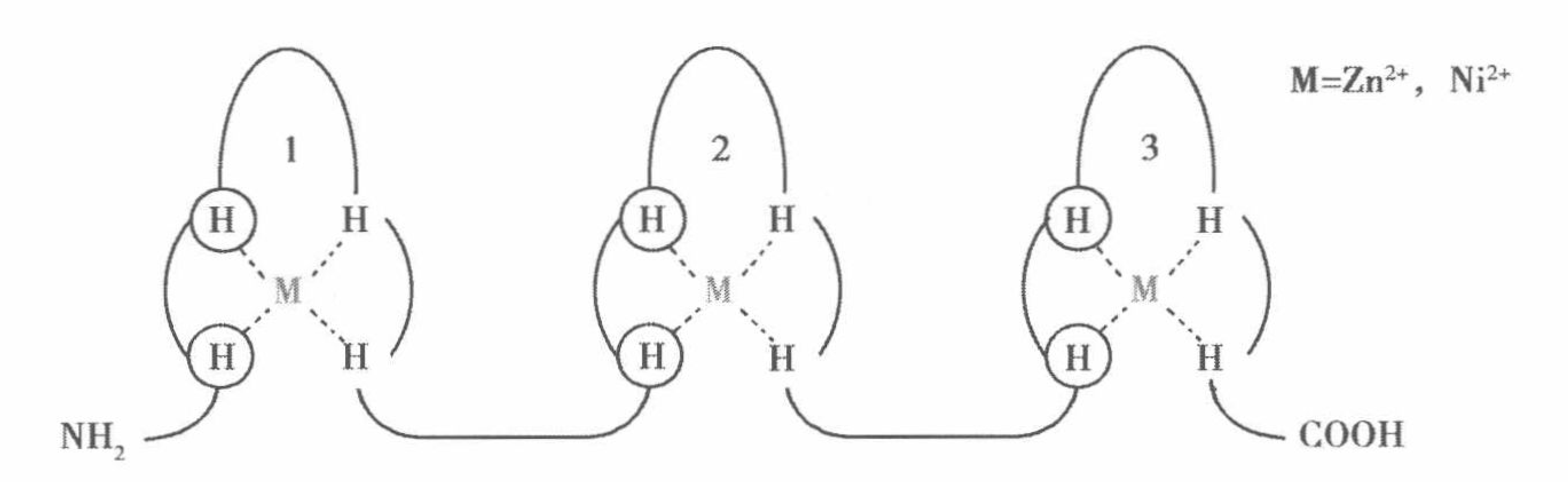


图 5.4 H_4Sp1 的框架图。取代 Cys→His 的位置用小圆圈标出。氨基酸残基用单字符标示。

5.3.4.2 $(His)_4$ 型锌指结构域的多肽合成

合成 3 锌指 Sp1 之中指的 $(His)_4$ 突变多肽（H4Sp1f2），在 Rink 氨基树脂上用 9-芴甲氧羰基固相合成。肽链在 Shimadzu PSSM-8 合成仪上采用苯并三唑-1-基氧 3 四氢吡咯磷、六氟合磷酸、（PyBOP）-1-羟基苯并三氮唑、（HOBt）-*N*-甲基吗磷啉、（NMM）耦合系统标准工艺合成。被保护起来的 H4Sp1f2 肽树脂，用三氟乙酸和乙醇（95∶5）在室温下处理 2h，接着用高效液相色谱在 μBondesphere4C4-300（19×50mm）柱上纯化。产物的可靠性用飞行时间质谱（Kratos Kompact MALDI4）验证。

5.3.4.3 H_4Sp1 的构建

$(His)_4$ 突变体 H_4Sp1 通过将 Sp1 的 3 锌指结构域的 6 个半胱氨酸突变为组氨酸构建。编码 $(His)_4$ 的 DNA 片段，从 pUC-Sp1（530-623）用基于点突变的聚合酶链反应（PCR）得到，并且用 GeneRapid DNA（Amersham Biosciences）测序仪验证。扩增的片段用 *Bam*HI 和 *Eco*RI 切割后，插入简单降解过的 pEY-3b 中（见 5.3.1.1；参考文献［12］）。

5.3.5 AT 识别锌指蛋白

识别 AT 的锌指的设计和构建分别在 5.3.5.1 和 5.3.5.2 节中描述。

5.3.5.1 蛋白质设计策略

$(Cys)_2$ $(His)_2$ 型锌指蛋白 Sp1 中的 3 锌指结合于富含 GC 的序列，而 6 $(Cys)_2$ $(His)_2$ 型锌指蛋白 CF2-II 中的 3 锌指（4～6 指）结合于富含 AT 的序列。新的与 AT 富集序列结合的锌指蛋白，用交换这两个类型的锌指蛋白的螺旋来合成（图 5.5）。

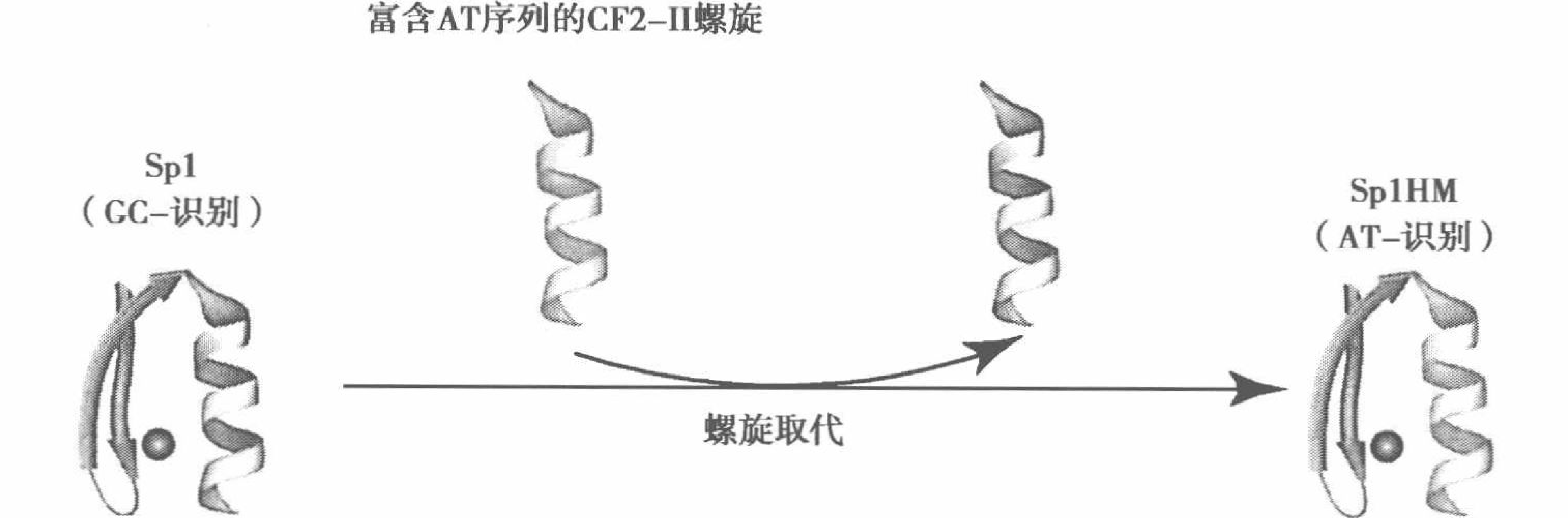


图 5.5 为交替识别 GC 和 AC 的螺旋取代策略示意图（见注 6 和注 7）。

5.3.5.2　Sp1HM 的构建

制备编码 CF2-II 的 α 螺旋的寡核苷酸。基于 PCR 的点突变，用 pUC-Sp1（530～623）作为模板来进行。PCR 扩增的突变片段用 *Bam*HI 和 *Eco*RI 降解后，插入简单地降解过的 pEV-3b 质粒中（见 5.3.1.1；参考文献 [13]）。

5.4　注

1. Krüppel 型连接子（Thr-Gly-Glu-Lys-Pro，TGEKP）在许多锌指蛋白中是保守的，因而被选来连接 Sp1 的锌指结构域。这些人工多锌指蛋白显示出扩展了的序列特异性，并且它们对序列的偏好取决于模体数量和 Sp1 3 锌指 DNA 结合结构域的特征。

2. 显然，足迹分析证明，现有的人工锌指蛋白中 Sp1ZF6 和 Sp1ZF9，使用它们所有的锌指结构域分别与 DNA 序列中的最少 18 和 27 个连续的 GC 富集碱基对结合。

3. 新设计的 6 锌指蛋白 Sp1ZF6（Gly)$_7$ 和 Sp1ZF6（Gly)$_{10}$ 能够在两个末端的结合部位交界处引起 DNA 弯折，并且，两个 3 锌指模体间连接子的长度对整个 DNA 弯折的方向起关键作用。相检测强烈地支持所引起的 DNA 弯折朝向 DNA 的主沟槽，并且，Sp1ZF6（Gly)$_7$ 引起 DNA 弯折最强烈的方向改变。

4. 通过用不同的带电连接子连接 Sp1 的两个 DNA 结合结构域，我们也制造了 6 锌指蛋白 Sp1ZF6（Gly · Arg)$_4$ 和 Sp1ZF6（Gly · Glu)$_4$（见参考文献 [14]）。

5. 新的（His)$_4$ 型锌指蛋白在自然界中从未观察到。特别有趣的是，人工（His)$_4$ 型 Sp1 的锌指结构域折叠和结合于 DNA 都类似于野生（Cys)$_2$（His)$_2$ 型 Sp1。

6. 在大多数天然蛋白中，含（Cys)$_2$（His)$_2$ 型锌指结构域的，都结合到 GC 富集序列，AT 识别的锌指蛋白非常稀少。用 α 螺旋替换，制造识别 AT 富集序列的锌指蛋白，是方便的和有用的。

7. 我们的 α 螺旋替换策略，在制造与 AT 富集和 GC 富集交替序列结合的锌指蛋白也是有效的。

参考文献

1. Nagaoka, M. and Sugiura, Y. (2000) Artificial zinc finger peptides: creation, DNA recognition, and gene regulation. *J. Inorg. Biochem.* 82, 57-63.

2. Imanishi, M., Hori, Y., Nagaoka, M., and Sugiura, Y. (2001) Design of novel zinc finger proteins: towards artificial control of specific gene expression. *Eur. J. Pharm. Sci.* 13, 91-97.

3. Pabo, C. O., Peisach, E., and Grant, R. A. (2001) Design and selection of novel Cys_2His_2 zinc finger proteins. *Annu. Rev. Biochem.* 70, 313-340.

4. Pavletich, N. P. and Pabo, C. O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1Å. *Science* 252, 809-816.

5. Pavletich, N. P. and Pabo, C. O. (1993) Crystal structure of a five-finger GLIDNA complex: new perspectives on zinc fingers. *Science* 261, 1701-1707.

6. Kim, C. A. and Berg, J. M. (1996) A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat. Struct. Biol.* 3, 940-945.

7. Nolte, R. T., Conlin, R. M., Harrison, S. C., and Brown, R. S. (1998) Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. USA* 95, 2938-2943.

8. Wütrich, K. (1986) *NMR of Proteins and Nuleic Acids*. Wiley Interscience, New York, NY.

9. Kuwahara, J. and Coleman, J. E. (1990) Role of the zinc (II) ions in the structure of the three-finger DNA binding domain of the Sp 1 transcription factor. *Biochemistry* 29, 8627-8631.

10. Kamiuchi, T., Abe, E., Imanishi, M., Kaji, T., Nagaoka, M., and Sugiura, Y. (1998) Artificial nine zinc-finger peptide with 30-base-pair binding sites. *Biochemistry* 37, 13827-13834.

11. Imanishi, M., Hori, Y., Nagaoka, M., and Sugiura, Y. (2000) DNA-bending finger: artificial design of 6-zinc finger peptides with polyglycine linker and induction of DNA bending. *Biochemistry* 39, 4383-4390.

12. Hori, Y., Suzuki, K., Okuno, Y., Nagaoka, M., Futaki, S., and Sugiura, Y. (2000) Artificial zinc finger peptide containing a novel His_4 domain. *J. Am. Chem. Soc.* 122, 7648-7653.

13. Nagaoka, M., Doi, Y., Kuwahara, J., and Sugiura, Y. (2002) Novel strategy for the design of a new zinc finger: creation of a zinc finger for the AT-rich sequence by α-helix substitution. *J. Am. Chem. Soc.* 124, 6526-6527.

14. Imanishi, M. and Sugiura, Y. (2002) Artificial DNA-bending six-zinc finger peptides with different charged linkers: Distinct kinetic properties of DNA binding. *Biochemistry* 41, 1328-1334.

6 独体——基于纤连蛋白类型Ⅲ结构域框架的抗体模拟

Akiko Koide and Shohei Koide

概要 我们发展了用人纤连蛋白的第10个纤连蛋白类型Ⅲ结构域（FNfn10）作为框架以显示用于与其他分子结合的多重表面链套的用法。我们把具有新的结合功能的FNfn10变体称为“独体（孤体、单小体）”（FNfn10是一个由94个氨基酸组成的小蛋白质，它具有近似免疫球蛋白折叠的β三明治结构）。它是高度稳定的，没有二硫键或金属离子，它在细菌中高水平地表达为正确的折叠形式。这些期望的物理性质，使得FNfn10框架与其他任何展示技术，事实上都可以比较。本章描述实物库构建和筛选以及独体的生产。
关键词 组合（实物）库；噬菌体展示；酵母双杂交系统；过表达

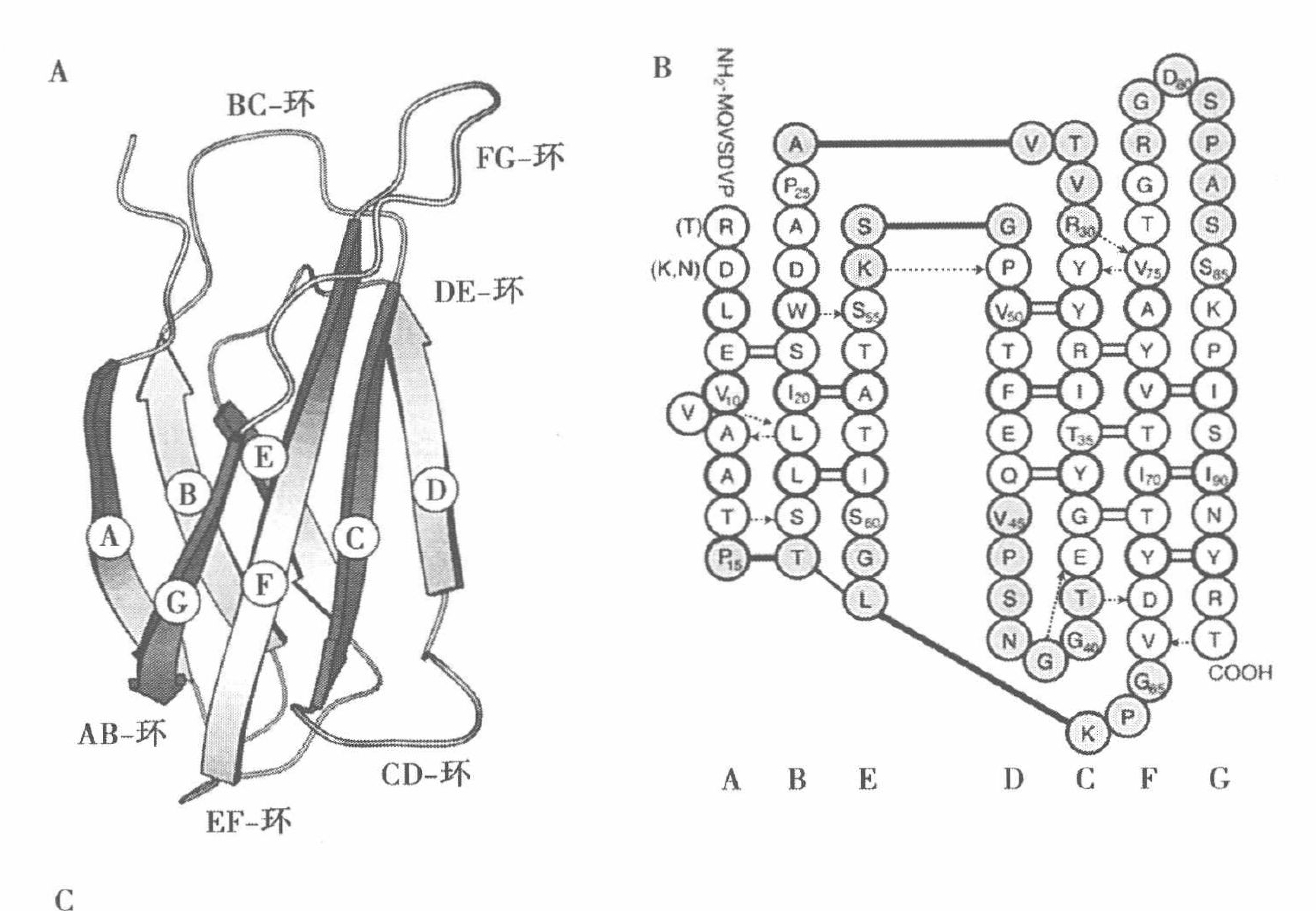
6.1 介　　绍

蛋白质工程的一个主要目的是生产能够与特定靶分子结合的新的生物分子。这样的分子会应用于生物技术的许多领域，包括诊断、治疗和催化。虽然存在大量的天然蛋白，可供这些应用和开发，但用工程手段制造新的、有用的生物分子的办法仍是有吸引力的。因为，它不受天然生物分子可用性的限制。抗体的基本构架中可变的链套呈现在基本不变的框架上，以及抗体工程的成功启发了一些研究组去发展“分子框架”[1,2]。这些努力的主要目标一直是在保留呈递多重肽片段以与靶分子结合的能力的同时，克服抗体和抗体片段的缺点，包括大尺寸、杂二聚特性（除单免疫球蛋白结构域外）、低构象稳定性和难于大规模生产。至今已发展出来的，这样的人工分子都是基于免疫球蛋白片段或者小单体蛋白。

我们发展了以纤连蛋白Ⅲ型结构域（FN3）为分子框架的应用。纤连蛋白是一个大蛋白，在胞外基质生成和细胞间相互作用中起基本的作用。它由多次重复的3类（Ⅰ、Ⅱ和Ⅲ）小结构域组成[3]。FN3是广泛存在的结构域，在所有动物蛋白中，其出现率估计为2%[4]。在人纤连蛋白中，有15个FN3的重复单位。我们的系统建立在FNfn10之上。FNfn10是小的（94个残基）单体，且不含二硫键。正确折叠的FNfn10在细菌中的高表达是很容易的[5]。FNfn10的三维结构[6,7]最好描述为β三明治，类似于抗体重链结构域的易变结构域，只是FNfn10有7条β链，而不是9条（图6.1）。FNfn10在其两端各含3条表面链套，具有载呈多重肽片段的能力。FNfn10的这些特性使其成为分子框架的理想候选者。

我们证明了通过筛选离散化表面链套残基的FNfn10组合库，可以设计制造出新结合蛋白（称为独体），得到的独体，保留框架蛋白的整体结构和良好的构象稳定性[8]。此外，因为FNfn10不需要二硫键来协助恰当的折叠和维持稳定性，所以独体兼容于任何筛选方法，包括噬菌体展示[8]，肽-核糖核酸融合[9]和体内方法，如酵母双杂交系统[10]。我们还证明了独体可以很容易地用于细胞内[10]。

我们在本章描述构建组合库、用噬菌体展示和酵母双杂交技术筛选蛋白库和用细菌生产独体的方法。



C

```
        BamHI NdeI
  1 GGATCCCATATGCGTGTTTCTGATGTTCCGCGTGACCTGGAAGTTGTTGCTGCGACCCCG
  1 GlySerHisMetArgValSerAspValProArgAspLeuGluValValAlaAlaThrPro

                BclI PvuII         PstI                       BsiWI
 61 ACTAGCCTGCTGATCAGCTGGGATGCTCCTGCAGTTACCGTGCGTTATTACCGTATCACG
 21 ThrSerLeuLeuIleSerTrpAspAlaProAlaValThrValArgTyrTyrArgIleThr

                                       EcoRI
121 TACGGTGAAACCGGTGGTAACTCCCCGGTTCAGGAATTCACTGTACCTGGTTCCAAGTCT
 41 TyrGlyGluThrGlyGlyAsnSerProValGlnGluPheThrValProGlySerLysSer

                                     SalI          Bst1107I
181 ACTGCTACCATCAGCGGCCTGAAACCGGGTGTCGACTATACCATCACTGTATACGCTGTT
 61 ThrAlaThrIleSerGlyLeuLysProGlyValAspTyrThrIleThrValTyrAlaVal

                               SacI
241 ACTGGCCGTGGTGACAGCCCAGCGAGCTCCAAGCCAATCTCGATTAACTACCGTACCTAG
 81 ThrGlyArgGlyAspSerProAlaSerSerLysProIleSerIleAsnTyrArgThr•••

         XhoI BamHI
301 TAACTCGAGGATCC
```

图 6.1 A. FNfn10 结构的框架图。显示从 A 到 G 的 β 链和连接它们的链套。B. FNfn10 的氨基酸序列及其 β 链拓扑。链套中残基有灰色背景。结构中的氢键用箭头和线标示。C. 我们研究中使用的一个 FNfn10 合成基因的核酸序列。

6.2 材　　料

6.2.1 培养基

(1) Luria-Bertani (LB)：每升水加 10g 菌用胰蛋白胨、5g 酵母抽提物和 10g 氯化钠，pH7.0。用平板时，加 15g/L 菌用琼脂糖。

(2) SOC：将 20g 胰蛋白胨、5g 酵母抽提物和 0.5g 氯化钠溶于约 900ml 水中；加入 10ml 250mmol/L 氯化钾和 5ml 2mol/L 氯化镁；并将最终体积调至 1L。高压灭菌、冷却，并加入 20ml 1mol/L 葡萄糖，用 0.2μm 滤膜过滤除菌。

(3) 2xYT：每升水加入 16g 菌用胰蛋白胨、10g 酵母抽提物，pH7.0。

(4) Superbroth（超级肉汤）：每升水加入 12g 菌用胰蛋白胨、24g 酵母抽提物和 5g 甘油。高压灭菌后加 50ml 含有 0.17mol/L 磷酸二氢钾和 0.72mol/L 磷酸氢二钾的灭菌溶液。

(5) M9-胰蛋白胨：将 6.4g 7 水合磷酸氢二钠、1.5g 磷酸二氢钾、0.25g 氯化钠、0.5g 氯化铵和 5g 菌用胰蛋白胨在 500ml 水中混合，灭菌。冷却至 50°C，并加入 5ml 50%葡萄糖、1ml 1mol/L 硫酸镁和 50μl 1mol/L 氯化钙。

(6) 抗生素最终浓度：100μg/ml 氨苄青霉素（Ap）；30μg/ml 氯霉素和 10μg/ml 四环素（Tc）。

(7) 基本硫酸卡那霉素（Km）板：将 6g 菌用琼脂糖混入 320ml 水中并灭菌，然后冷却至 50°C。在另外一瓶中加入 4.2g 磷酸氢二钾、1.8g 磷酸二氢钾、0.4g 硫酸铵、0.2g 柠檬酸钠和 64ml 水，灭菌后冷却至 50°C。加入 64ml 培养基到这个瓶中，4ml 10mg/ml 尿嘧啶、1.6ml 50%葡萄糖、0.32ml 1mol/L 硫酸镁、0.66ml 30mg/ml Km 和 20μl 100mg/ml 硫胺。

(8) 酵母抽提物-蛋白胨-葡萄糖（YPD）：20g 蛋白胨、10g 酵母抽提物和 20g 葡萄糖溶于 1L 水中。用平板时，加入 20g/L 菌用琼脂糖。

(9) 酵母全价（YC）葡萄糖 leu-his-trp-ura-：为得到 1L 试剂，混合 1.2g 的无氨基酸无硫酸铵酵母氮源（Difco），5g 硫酸铵、10g 琥珀酸、6g 氢氧化钠、0.6g 全价补充混合物-leu-his-trp-ura-（Q-BIOgene）、0.1g 半胱氨酸和 960ml 水，并灭菌。加入 40ml 灭菌的 50%葡萄糖并彻底混合。用平板时，加入 20g/L 菌用琼脂糖。

(10) YC Glc his-ura-trp-：灭菌前，加入 0.1g 亮氨酸到 YC Glc leu-his-trp-ura-培养基中。

(11) YC Glc trp-：灭菌前，加入 0.1g 亮氨酸、0.05g 组氨酸和 0.1g 尿嘧啶到 YC Glc leu-his-trp-ura-培养基中。

(12) YC Glc his-ura-：灭菌前，加入 0.1g 亮氨酸到 YC Glc leu-his-trp-ura-培养基中，并用 910ml 水而不是 960ml 水。灭菌后，加入 50ml 灭菌的 2mg/ml 色氨酸溶液。

(13) YC 半乳糖蜜三糖 leu-his-trp-ura-：在 YC Glc leu-his-trp-ura-培养基中用 100ml 灭菌 20%半乳糖和 50ml 灭菌 20%蜜三糖代替葡萄糖。用 850ml 水而不是 960ml。

(14) YC Gal Raf his-trp-ura-：灭菌前，加入 0.1g 亮氨酸到 YC Gal Raf leu-his-trp-ura-培养基中。

6.2.2　试剂

(1) 聚乙二醇（PEG）/氯化钠溶液：混合 66.6g PEG8000，78g 氯化钠和 317ml 水并灭菌。

(2) HEPES 缓冲液：1mmol/L HEPES，pH7.0（HEPES：*N*2-羟乙基哌嗪-*N*′-2-

乙烷-磺酸）。

（3）涂布溶液：20μg/ml 配体溶于 0.1mol/L 碳酸氢钠，pH8.6。

（4）Tris 缓冲盐水（TBS）：50mmol/L Tris-氯化氢和 150mmol/L 氯化钠，pH7.5。

（5）TBS-Tween-20（TBST）：TBS 和 0.5%（*V/V*）Tween-20。

（6）5%灭菌牛血清蛋白（BSA）：过滤除菌的 BSA 于水中并储藏在 4°C。

（7）TBST-BSA：TBST 加入 1mg/ml BSA。

（8）酸性洗脱缓冲液：0.1mol 氯化氢，加入甘氨酸和 1mg/ml BSA，酸碱度调至 pH2.2。

（9）2mol/L Tris-碱。

（10）琼脂糖-磷酸溶液：每板混合 8ml 水、2ml 0.5mol/L 磷酸钾缓冲液，pH7.0，0.07g 琼脂糖（根据平板的数量，加倍用量）。

（11）2×β-半乳糖苷酶检验溶液：120mmol/L 磷酸氢二钠、80mmol/L 磷酸二氢钠、20mmol/L 氯化钾、2mmol/L 硫酸镁、0.54%β-巯基乙醇、0.0008%十二烷基硫酸钠、8mg/ml 2-硝基苯-β-D-半乳糖苷，pH7.0。等分为小份存储于−20°C。

（12）缓冲液 B：20mmol/L Tris-氯化氢，pH8.0，500mmol/L 氯化钠。

（13）缓冲液 C：20mmol/L Tris-氯化氢，pH8.0，500mmol/L 氯化钠和 500mmol/L 咪唑。

（14）缓冲液 D：50mmol/L 磷酸钠和 500mmol/L 氯化钠，pH8.0。

（15）Tris-氯化氢-EDTA（TE）缓冲液：10mmol/L Tris-氯化氢和 1mmol/L EDTA，pH8.0。

（16）50×Tris-乙酸盐-EDTA 缓冲液：溶解 242g Tris-碱、57.1ml 冰乙酸和 18.6g EDTA 于重蒸水中至终体积 1L。调节 pH 至 8.3。

（17）大肠杆菌菌株：CJ236（Bio-Rad）、KC8（Origene）、XL-1 Blue（Stratagene）和 SS-320（见参考文献［11］；注 1）。

（18）酵母菌株：EGY48 和 RFY206（Origene；参考文献［12］和［13］）。

6.2.3 载体

（1）pAS38：融合于 *M*13 基因 III 的 C 端结构域的编码 *FNfn*10 的 pBluescript SK（+）噬粒（Stratagene）的衍生物。我们用大肠杆菌偏好密码子构建了 *FNfn*10 基因[8]。这是为了独体的单价展示。此载体的 *Pvu*II 片段在与 pBluescript SK（+）的相反方向。

（2）pAS45：从 pET15b（Novagen）衍生的独体表达质粒。它表达带有 His 标签的独体融合蛋白。独体基因含有一个 Arg6 到 Thr 的突变，为了消除凝血酶的一个酶切位点，从一开始就已引入[8]。

（3）pAS47：在 FG-链套区含有一终止密码子的 pAS38 的衍生物（图 6.1A）。这是为了在构建 FG-链套区残基离散化的实物库时用作突变模板。

（4）JCFN：编码独体-基因 VIII 融合蛋白的 JC-M13-88[14] 的衍生物。这是为独体多价展示用的 M13 噬菌体体系。

（5）pYT45：编码 B42（激活蛋白）-独体融合蛋白的 pYesTrp2（Invitrogen）的

衍生物[10,15]。这是用来构建酵母双杂交筛选用的独体库（见注 2）。

（6）pSH18-34：酵母双杂交筛选用的 LacZ 报告质粒（Origine）。

（7）pEG202：酵母双杂交筛选用的 LexA-诱饵质粒（Origine；见注 2）。

独体质粒和实物库可从通信作者处得到。

6.3 方　　法

6.3.1 实物库构建

我们用 Kunkel 突变技术构建大多数的实物库[16]。虽然我们证明了 AB 链套可用来结合目标分子，但多数情况下，我们用 FNfn10 的“上”端（BC、DE 和 FG 链套；图 6.1A）作为目标分子结合位点[10]。我们的方法基于 Bio-Rad 的 Mutagene kit[17]。用大肠杆菌 SS320 做电转化，可以很容易地制备含有大约 10^9 个独立克隆的实物库。

6.3.1.1 尿嘧啶单链噬粒的制备（用于单价噬菌体展示和酵母双杂交筛选）

（1）用 pAS47、pAS38（用于独体-基因 III）或 pYT45（用于 B42-独体）转化 CJ236，在 LB-Ap 平板中选择菌落。

（2）把一菌落接种到 2ml 2×YT-Ap-氯霉素培养基，37℃摇动过夜。

（3）将 30ml 补充了 0.25μg/ml 尿苷的预热 2×YT-Ap 培养基装入 250ml 规格带凹陷三角烧瓶，用 300μl CJ236 过夜培养液接种，再加入 10^{10} 个/ml 辅助噬菌体 KO7（Promega）。在强力摇动下于 37℃生长 2h。加入 Km 使终浓度达到 70μg/ml。在强力摇动下于 37℃生长过夜。

（4）于 4℃以 12 000*g* 离心［10 000r/min，用 SS-34（Sorvall）或等效的转子］10min。将上清液移入新试管，再次以 12 000*g* 离心 10min。将上清液移入新试管。加入 4.5ml PEG/NaCl 溶液，并彻底混合。4℃保育过夜。

（5）于 4℃以 17 200*g* 离心（12 000r/min，SS-34 转子）20min。抛弃上清液将试管倒放在一叠纸巾上 1min。用纸巾擦去残余液体。

（6）将噬粒沉淀悬浮于 1ml TBS 中，并移入微离心管。在微型离心机上以最大速度离心 2min。将上清液移入新微离心管。加入 150μl PEG/NaCl 溶液，彻底混合。在冰上保育 30min。

（7）在 4℃以最大速度离心 10min。抛弃上清液，稍加离心，并用移液器除去所有液体。将沉淀悬浮于 1ml TBS 中。

（8）用 XL-1 Blue 和 CJ236 测定噬粒浓度。将 100μl 新鲜大肠杆菌（600nm 光密度 OD 值为 0.5～1.0）和 10μl 系列稀释噬粒溶液混合。室温保育 15min，并铺在 LB-Ap 板上。在 37℃保育过夜。用 CJ236 的滴定值应该是 10^{12}～10^{14} 个/ml，并且应该比用 XL-1 Blue 的值大 10^4。

（9）用 Qiagen 单链 DNA 制备试剂盒制备尿嘧啶单链 DNA。

6.3.1.2 尿嘧啶单链噬菌体的制备（用于多价噬菌体展示载体，JCFN）

（1）用 JCFN 质粒转染 CJ236。热激之后，与 30ml 预热的 2×YT 和 500μl 新鲜的

CJ236（OD_{600} 值 0.5～1.0）混合，再于 37℃摇动 6h。

（2）完成 6.3.1.1 节的步骤（4）至（9）。为测定浓度，混合 3ml LB 和 0.7%琼脂糖（熔化后冷却至 45℃），10μl 连续稀释的噬粒溶液和 100μl 新鲜的大肠杆菌。混匀并立即冲入 LB 平板中。在 37℃培养保育过夜并测定滴度。

6.3.1.3 突变寡核苷酸的制备

（1）设计一寡核苷酸，取决于 GC 含量，使其分别含有大约 20 和 15 个碱基的 5′或 3′互补区。我们使用密码 NNK 和 NNS 来进行“硬”随机化，其中 N 代表所有核苷酸的等摩尔混合；K 代表 G 和 T 的等摩尔混合；S 代表 G 和 C 的等摩尔混合。

（2）将约 10μg 溶解于水的粗寡核苷酸做丙烯酰胺凝胶电泳，切出对应于正确分子质量的带，将寡核苷酸用电洗脱（见注 3）。

（3）将 2μg 纯化的寡核苷酸、3μl T4 多聚核苷酸激酶缓冲液、5U T4 多聚核苷酸激酶（New England Biolabs）、1.5μl 10mmol/L 腺苷三磷酸混合，加水至 30μl。37℃保育 2h，然后 65℃保育 15min。将溶液在－20℃保存。

6.3.1.4 双链 DNA 合成

（1）混合大约 1μg 尿嘧啶单链 DNA、2μl 磷酸化的寡核苷酸（见 6.3.1.3），1μl 10×退火缓冲液（Bio-Rad Metagene kit），加水使体积达 10μl（见注 4）。制备一无寡核苷酸的对照反应。

（2）使用聚合酶链反应（PCR）仪，在 70℃保温 5min，然后以每分钟降 1℃的速度，降温至 30℃。将试管置于冰上。

（3）在退火混合物中，加入 1μl 10×合成缓冲液（Bio-Rad Metagene kit）、1μl T4 DNA 连接酶和 1μl T4 DNA 聚合酶（New England Biolabs）。在冰上和室温相继保温各 5min，37℃保温 30min，75℃保温 15min，然后冷却至室温。将 1μl 样品通过琼脂糖胶（电泳）。含有寡核苷酸的样品会得到比无核苷酸的对照样品所得分子质量高的产物。

（4）把一滴混合物放在飘在水上的 0.025μm 透析膜碟上，透析合成混合物 1h，然后在新的微离心管中复原透析过的混合物。使用前保存在冰上。

6.3.1.5 电穿孔胜任细胞的制备

（1）在 350ml 含 Tc 的超级肉汤中培养大肠杆菌 SS-320 直至 OD_{600} 达到 0.8。

（2）在冰浴中快速冷却三角瓶中的培养液，然后放置在冰上 15min。将培养液移至离心瓶，在 0～2℃以 3900*g* [4700r/min，用 JA-10（Beckman）或等效转子] 离心 5min（见注 5）。

（3）撇去上清液，通过在冰上敲击离心管，将细胞悬浮于 10ml HEPES 缓冲液中。加入 390ml HEPES 缓冲液，并搅动使细胞悬浮。在 0～2℃以 3900*g* 离心 5min。

（4）撇去上清液。将细胞悬浮于 10ml HEPES 缓冲液中并将悬浮物移至 50ml 离心管。用缓冲液漂洗第一个瓶子并在 50ml 离心管中恢复细胞悬浮。

（5）在 0～2℃以 4300*g*（6000r/min，用 SS-34 转子）离心 10min。去上清液。在冰上敲击离心管使细胞悬浮于 10ml 10%甘油中，然后加入 20ml 10%甘油，并将离心管颠倒数次以使悬浮物混合。

(6) 在 0～2℃以 4300g 离心 10min。去上清液。加 300μl 10%甘油，并在冰上敲击离心管使细胞悬浮。检查体积，加 10%甘油使终体积达 700μl。将悬浮物等分至两个离心管中（各 350μl）。

6.3.1.6　噬粒 DNA 电穿孔转染

(1) 将冷却的 DNA［见 6.3.1.4，第（4）步］加入放在微离心管中的 350μl 电转化感受态细胞中，并在冰上保温 5min。将混合物移至预冷的电穿孔试管中（2mm 间隙）。

(2) 电穿孔细胞，如在 2.5kV 高压模式下使用 BTX ECM395 电穿孔器。

(3) 立即在试管中加入 1ml SOC，并使细胞悬浮。转移细胞至 250ml 三角瓶。加 1ml SOC 到试管中，清洗残留细胞并移入三角瓶中。再重复此操作 3 次。

(4) 三角瓶中加入 20ml SOC。在 37℃以 180r/min 摇动三角瓶 30min。并通过放置部分稀释的悬浮物至 LB-Ap 平板检查浓度。

(5) 将全部悬浮物加至 500ml 预热的 2×YT-Ap 中。为酵母双杂交的质粒制备，在 37℃摇动过夜并制备质粒。为噬粒制备，加入 10^{10} pfu/ml 辅助噬菌体 KO7，在 37℃摇动过夜，并如 6.3.1.1 节所述制备噬粒。

为噬菌体 DNA 的转化，在第（3）步之后，加入 1.5ml 中对数期 SS-320 细胞，然后将悬浮物移至预热的 120ml 2×YT-Tc。按照 6.3.1.2 节所述测定噬菌体浓度。37℃保育 4h。按照 6.3.1.1 节所述制备噬菌体。

6.3.1.7　酵母实物库构建

为酵母双杂交筛选，在大肠杆菌中，我们按照 6.3.1 节的描述构建了载体库，然后将这个库引入酵母中。酵母转化系基于 Gietz 的方法[18]。用这个方法，我们在典型情况下得到约 10^6 个独立克隆。

(1) 在 30℃，20ml YPD 中摇动过夜生长酵母 EGY48。将培养液加入预热的 300ml YPD 使 OD_{600} 值为 0.25。让细胞生长直至 OD_{600} 值接近 1。

(2) 在室温下以 3600g（4500r/min，JA-10 转子）离心 5min。撇去上清液，并悬浮细胞于 300ml 灭菌水中。重复离心和悬浮。再离心，悬浮细胞于 25ml 灭菌水，并移至 50ml 离心管中。室温以 3000g（5000r/min，SS34 转子）5min。撇去上清液，并悬浮细胞于 1.2ml 灭菌水。

(3) 将 35μl 细胞悬浮物放在一边用作负对照。在余下的细胞悬浮物中加入 7.2ml 50% PEG3350、1.08ml 1mol 乙酸锂、500μl 携带子单链 DNA（Origene）、45μg DNA 和 900μl 水。彻底混匀，并每份 360μl 分装入 30 个微离心管中。在负对照管中加入 1/30 除 DNA 外的上述材料（即 PEG、乙酸锂和携带子 DNA），彻底混匀。

(4) 在 42℃保育 40min。在微离心机上以 10 000r/min 转速室温离心 10s，撇去上清液。悬浮每管中的细胞于 150μl 水中。将细胞铺在直径为 15cm 的筛选平板上（YC Glc trp-）上。30℃保育 3d。

(5) 每皿加入 15ml 水，用灭菌的盖玻片刮散所有菌斑。把所有细胞悬浮在 500ml 离心管中，并以 3900g（4700r/min，JA-10 转子）离心 5min。撇除上清液，悬浮细胞

于 500ml 水中。再次离心并撇除上清液。悬浮细胞于 50ml 水中，并测量 OD_{600} 以确定细胞密度（1OD_{600} 对应于约 2×10^7 个/ml）。加入甘油至终浓度 15%（*V*/*V*）。分装并保存细胞悬浮物于－80℃。

6.3.2 噬菌体展示库分类

6.3.2.1 扫视（panning）

（1）加入 50μl 涂渍溶液到酶联免疫吸附试验酶标板（Nunc Maxisorp，可分小孔，C-bottom）的小孔中（每孔 1μg 的配体）。把酶标板放入一铺有湿纸巾的密封盒子中。保持盒子在 4℃过夜，或 37℃1h。

（2）用移液器除去液体，并用水清洗小孔两次。加入 360μl 含 3%BSA 的 TBS 于小孔中。37℃保育 1h。为了下一轮实验，制备一对照小孔；加入无配体的缓冲液，并用 BSA 封闭小孔。

（3）从小孔移除封闭液，用水清洗两次。加入 12μl 5%BSA 溶液和 50μl 噬菌体悬浮物（10^{11} 个/孔）。室温保育 2h。

（4）移除噬菌体溶液，然后加入 360μl TBST。用吸管上下有力地吸动。等待 5min 然后移除溶液（第一轮一次，以后每轮 5 次）。

（5）在小孔中加入 50μl 含有 10μmol/L 配体的 TBST 溶液。37℃保育 1h，用吸管上下有力地搅动，回收洗脱物。也可以加 50μl 酸性洗脱缓冲液到小孔中，室温等待 10min。用吸管上下有力地搅动。回收第二次洗脱物到装有 3μl 2mol/L Tris-碱的微离心管中，准备中和。

（6）混合 5ml 新鲜的 XL-1 Blue（OD_{600} 值为 0.5～1；在 LB-Tc 中生长的）和洗脱的噬菌体。对噬菌体，加入 100ml 2×YT [和异丙基巯基半乳糖苷（IPTG）直至 1mmol/L 以得到独体的高产出]，并在剧烈摇动的同时，于 37℃保育 6h（更长的保育时间可能导致独体基因删除）。加入 2×YT 后立即按 6.3.1.2 节所描述的测定浓度。对噬粒，混合 5ml 新鲜的 XL-1 Blue 和洗脱的噬粒，37℃保育 20min、加入 100ml 2×YT（以及 IPTG，如果需要）、100μg/ml AP 以及 10^{10} 个/ml 的辅助噬菌体 KO7；在 37℃保育 2h 并保持有力地摇动。Km 至终浓度 70μg/ml 并保育过夜。保育 20min 后立即按 6.3.1.1 节描述测定浓度。按 6.3.1.1 和 6.3.1.2 节所述，分离噬菌体（或噬粒）。重复整个步骤，直至洗脱的噬菌体（或噬粒）数增加。

6.3.2.2 单噬菌体克隆的鉴定

（1）从单克隆中制备噬菌体（或噬粒）。对噬粒，在 2ml LB-Ap 上过夜生长一单菌落。混合 20μl 培养液、2ml LB-Ap（和 IPTG，如愿意的话）和 10^{10} 个/ml 辅助噬菌体 KO7。37℃剧烈摇动保育过夜。对噬菌体克隆，用灭菌的牙签接触一单菌斑。将牙签放在 2ml 含 100μl 新鲜 XL-1 Blue 的 LB-Ap（和 IPTG，如果需要的话）中，37℃剧烈摇动保育 6h。按 6.3.1.1 和 6.3.1.2 节所述，制备噬菌体（或噬粒）。

（2）对每一个要测试的克隆，完成 6.3.2.1 节的第（1）和第（2）步，来准备一带有固定目标分子的小孔。同时用不含目标分子的涂渍溶液涂渍，来制备另一对照用小

孔（见注6）。

(3) 加入 5μl 5%BSA 溶液和 50μl 从单克隆制备的噬菌体悬浮液到配体涂渍的和对照的小孔中。视相互作用强弱，调节噬菌体浓度（典型值：每孔 10^8 ～ 10^{11} 个）。在搅拌器上室温保温 2h。

(4) 用 TBST 冲洗 5 次。用喷水瓶在所有小孔中加入 TBST，然后清除液体，将平板翻转面朝下拍在一叠纸巾上。

(5) 在每一个孔中，加入 50μl 抗噬菌体-辣根过氧化物酶的抗体溶液（Phamacia；在 TBST∶BSA 中稀释 1∶2000），在搅拌器上室温保温 1h。用 TBST 清洗 5 次。

(6) 加入 50μl 1step turbo-TMB（Pierce），等待约 10min 直到蓝色显现，并加入 50μl 2mol/L 硫酸以终止反应（使用有滤嘴的枪尖以保护移液器）。用平板读出器测量 OD_{450}。选择对配体涂渍的小孔有高亲和力并对对照小孔呈低亲和力的克隆，以备进一步检测（测序、亲和力优化和蛋白质生产）。

6.3.3 酵母双杂交筛选

6.3.3.1 目标（“诱饵”）质粒的制备

在 pEG202 中克隆诱饵基因，使得诱饵与 LexA 在同一可读框表达。重要的是确认 LexA-诱饵自己不激活报告基因，也不与野生型 FNfn10 相互作用。用相互作用交配法，这很容易测试[13,19]。

(1) 用 LexA-诱饵质粒［或 pEG202 作负对照，pBait（Origene）作正对照］和 pSH18-34（β-半乳糖苷酶报告质粒；Origene）转化 RFY206，并在 YC Glc his-ura-平板上选择。也用 pYesTrp2（激活子 B42 质粒；Invitrogen）、pTarget（正对照；Origene）或 pYT45 转化 EGY48，并在 YC Glc trp-平板上选择。

(2) 在 YC Glc his-ura-平板上两个菌斑的每一个上用 LexA 质粒和报告质粒上平行划痕。在 YC Glc trp-平板用 B42 质粒以同样的方式划痕。30℃保育两块平板 1～2d。

(3) 复制两块平板至一 YPD 平板。从两块板上来的划痕应该互相垂直，以使 EGY48 细胞和 RFY206 细胞在交叠处杂交。30℃保育过夜。

(4) 将 YPD 平板复制至 YC Glc leu-his-trp-ura、YC Glc his-trp-ura-和 YC Gal Raf leu-his-trp-ura-平板，并在 30℃保育 3d。所有杂交细胞应该在 YC Glc his-trp-ura-平板上生长。在 YC Gal Raf leu-his-trp-ura-平板上的生长指示“诱饵”和“捕食者”的相互作用。其他平板用于对照。

6.3.3.2 酵母双杂交库筛选

含有诱饵的 RFY206 与含有独体库的 EGY48 杂交，并选出含有独体和诱饵的细胞。酵母杂交的应用，使得可以有效地筛选多个库[13]。

(1) 在 30℃，10ml YC Glc his-ura-培养基中，生长 16h 驻有 pSH18-34 和 LexA-诱饵质粒的 RFY206。

(2) 用 1.0 OD_{600} ＝ 2.0×10^7 个细胞/ml 换算，通过 OD_{600} 测定细胞浓度。10^8 个诱饵细胞放在微离心管中以 10 000r/min 的转速在微离心机上离心 30s，使细胞旋转沉降，用 200μl 培养基让细胞悬浮。

(3) 加入 10^7 个驻有独体的 EGY48 细胞到诱饵细胞中（见 6.3.1.7)，并在 YPD 平板上铺开。让细胞杂交 16h。

(4) 用 5ml（每次用 1ml）YPD 培养基将细胞从平板上冲洗下来。100 倍稀释后测量 OD_{600}，将 10^8 个细胞铺在 5 个 YC Gal Raf leu-his-trp-ura-培养基平板上。于 30℃保育这些平板 3～4d。

(5) 复制菌落到 YC Gal Raf leu-his-trp-ura-培养基平板和 YC Gal leu-his-trp-ura-培养基平板上。保育 16～24h。真的正克隆将只会在 YC Gal Raf leu-his-trp-ura-平板上生长。

(6) 在通风厨内，打开平板盖，加入足够氯仿以覆盖板表面。5min 后将氯仿倒入废物罐，并让剩余的氯仿挥发 10min[20]。

(7) 在微波炉内加热琼脂糖-磷酸盐溶液（每皿 10ml)，以溶解琼脂糖，在水浴中冷却到 50℃。在试管中加入 35μl 50mg/ml X-gal 溶液和 10ml 琼脂糖-磷酸盐溶液，搅拌，并倒入平板中。30℃保育。观察颜色。挑出呈现蓝色的克隆以备进一步检验。

6.3.3.3 分离独体的特异性检测

(1) 于 30℃，在 1.5ml YPD 培养基中，培养从库中筛选出来的感兴趣的酵母细胞 16h，并用 Y-DER 酵母 DNA 抽提试剂盒（Pierce）从酵母细胞中制备 DNA。

(2) 将透析后的酵母 DNA 用电穿孔注入大肠杆菌 KC8 细胞（见 6.3.1.4)，并在 LB-Ap 平板上挑选转化细胞。

(3) 在基本 trp-Km 平板上复制菌落，并于 37℃保育。选取两个菌落分别接种于 2ml LB-Ap 培养基中，于 37℃，培养 10h，并用标准方法抽提质粒 DNA。生长超过 10h 的 KC8 细胞常产生不纯的质粒。用标准的 DNA 测序法测定独体片段的序列（见注 7)。

(4) 用 B42-独体质粒转化 EGY48，并用相互作用杂交法测定其与各种诱饵的相互作用（见 6.3.3.1)。

6.3.3.4 液相 β-半乳糖苷酶测试

生化测量之前，诱饵与独体的亲和力可用液相 β-半乳糖苷酶测试半定量地加以确定。这个方法比 6.3.3.2 小节描述的平板测试更为定量。我们改进了这个方法使之适用于 96 孔平板格式（见注 8)。

(1) 用 pSH18-34、LexA-诱饵质粒（pEG202 的衍生物）和 B42 独体质粒 (pYesTrp2 衍生物）转化 RFY206 酵母菌株。把一个菌落接种至 2ml YC Glc his-ura-trp-，于 30℃摇动过夜。

(2) 稍稍离心以移除培养基，在温热的 YC Gal Raf his-ura-trp-中悬浮直至 OD_{600} 值为 0.2。于 30℃摇动 6h。测定 OD_{600} 值并以每 350μl 装入微离心管中。典型地，我们每次测量三遍。

(3) 将 2×β-半乳糖苷酶测试溶液与 Y-Per（Pierce）按 1∶1 混合作为工作溶液。加 350μl 工作溶液到每个样品中。多次翻转以混合。

(4) 30℃保育，同时检查颜色。当呈现黄色时，加 300μl 0.5mol/L 碳酸钠入离心管，并彻底混合。以最大速度离心微离心管 1min。测定上清液的 OD_{420} 值。

6.3.4 独体的克隆、表达、纯化和生物素化

当从库中筛选出具有所期望特性的独体后，其基因转染至大肠杆菌表达载体，独体则被表达纯化为一分离的蛋白质。我们在典型情况下每升培养液得到 10～15mg 独体。我们也描述了为在免疫化学应用中便于探测而使独体生物素化的操作步骤。

6.3.4.1 独体的克隆

用寡核苷酸 NdeMetThrFNF（5′-CGGGATCCCATATGCAGGTTTCTGATGT-TCCGACCGACCTGGAAGTTGTTGCTG-3′；它包含 Arg6 到 Thr 的突变）和 FNGKKGKR（5′-CCGACTCGAGTTACTATTTACCTTTTTTACCGGTACGGTAG TTAATCGAG-3′），扩增感兴趣的独体基因，然后用 *Nde*I 和 *Xho*I 降解并克隆到用同样的酶降解过的 pET15b（Novagen）中。通过测序确认表达载体中基因。

6.3.4.2 独体的表达

此操作步骤适合于 100ml 培养液，对初步鉴定而言，这样的量应该能产生足够的独体。

（1）用独体表达载体转化 BL21（DE3）细胞（Novagen）（见注 9）。

（2）接种转化细胞至 10ml M9-胰蛋白胨-Ap，并在强烈摇动下，于 30℃保育 10～16h。

（3）接种 2ml 预培养液至预热的 100ml M9-胰蛋白胨-Ap，并在强烈摇动下于 37℃保育。当 OD_{600} 值达到 0.7 时，加入 IPTG 至终浓度 0.5mmol/L。

（4）在强烈摇动下于 37℃保育 3h 多，并离心收集细胞。细胞可存储于－20℃，直至使用。

6.3.4.3 独体的纯化

此操作步骤适合于 100ml 培养液（见 6.3.4.2）。

（1）悬浮细胞沉淀于 6.4ml pH8.0 的 50mmol/L Tris-氯化氢中。

（2）加入 64μl 50mg/ml 溶菌酶和 128μl 50mmol/L PMSF。混合并于 37℃保育 15min。超声破碎并悬浮直至不再有高度黏滞性。

（3）加入 910μl 40mol/L 氯化钠。于 4℃以 27 000*g*（15 000r/min，SS34 转子）离心 10min。收集上清液。

（4）将蛋白质溶液放入装载了 0.1mol/L 氯化镍并用缓冲液 B 平衡过的 Hi-Trap 螯合柱中（1ml；Amersham-Pharmacia Biotech）。先用 6ml 缓冲液 B，再用 6ml 含 60mmol/L 咪唑的缓冲液 B 洗柱。

（5）用 6ml 缓冲液 C 洗脱独体。收集 0.5～1ml 每份的洗脱液到微离心管中。用十二烷基硫酸钠-聚丙烯酰胺凝胶电泳分析洗脱液。

6.3.4.4 独体的生物素化

我们连接生物素到 Lys 的氨基。这虽然不是位点特异的，但在 C 端的扩展部分成团的 Lys 应该是最易发生的修饰位点。这个方法可用来连接其他的片段，如荧光染料。

（1）完成 6.3.4.3 小节中纯化操作的步骤（1）～（4）。

（2）用 10ml 缓冲液 D 清洗柱子。

（3）溶解 1mg D-生物素-ε-氨基己酸羟基琥珀酰亚胺脂（Boehringer Mannheim）于 50μl 二甲基亚砜，并将其稀释至 3ml 缓冲液 D 中。注射 1ml 溶液到柱中并于室温黑暗中保温 1h。再重复反应 2 次。

（4）用 6ml 缓冲液 B 洗柱，并按 6.3.4.3 节第（5）步所述洗脱独体。

6.4 注

1. SS-320 不是商品，但按照 Sidhu 等的操作工艺[11]很容易合成。

2. 我们选择 pYesTrp2 作为捕食质粒（相对于诱饵而言——译者），主要是因为在这一质粒中存在允许 Kunkel 突变的 f1 起点。我们选择 pEG202 为诱饵质粒，因为该质粒在酵母中的维护较易，且较 pHybLex（Invitrogen）经济，后者需要 zeocin 抗生素。

3. 寡核苷酸的电洗脱。在一端用管夹封闭的透析管（MWCO 3500，18mm 宽）中，加入凝胶片和 1.8ml TE 缓冲液，然后在清除气泡后，用另一管夹封闭开口。将凝胶片放在透析管的一边，并淹没整个装置于一装满 1× Tris-乙酸盐-EDTA 缓冲液的凝胶盒中。用 200V 电压，电洗脱寡核苷酸 10min。从管中回收溶液。苯酚/氯仿抽提后，寡核苷酸可被乙醇沉淀回收。

4. 为同时离散化不同链套，一次反应可使用一个以上的寡核苷酸。

5. 为制备电穿孔感受态细胞，离心瓶、移液管和缓冲液用前必须冷藏。

6. 封闭前，对照孔必须用无目标分子的缓冲液涂覆。噬菌体对没有经受缓冲液涂覆步骤，只是简单地用 BSA 封闭的孔壁，只有显著低结合活性。

7. DNA 测序用模板，可从酵母菌落用 PCR 扩增来产生。这种情况下，在开始扩增循环之前，在 96℃加热 PCR 反应混合物 8min。

8. β-半乳糖苷酶测试可用 96 孔格式板实现。用深孔板（每孔 1.2ml），在半体积中完成反应。β-半乳糖苷酶反应停止和在一摇摆转子上以 2200g［3000r/min，用 JS-5.3（Beckman）或等效转子］离心 96 孔板 10min 之后，转移 300μl 上清液到 96 孔板，并用读板器测定 OD_{420}。

9. 为得到独体的高产率，使用新鲜转化的 BL21（DE3）细胞。

参考文献

1. Skerra, A. (2000) Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* 13, 167-187.

2. Nygren, P.-Å. and Uhlén, M. (1997) Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.* 7, 463-469.

3. Baron, M., Norman, D. G., and Campbell, I. D. (1991) Protein modules. *Trends Biochem. Sci.* 16, 13-17.

4. Bork, P. and Doolittle, R. F. (1992) Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA* 89, 8990-8994.

5. Plaxco, K. W., Spitzfaden, C., Campbell, I. D., and Dobson, C. M. (1996) Rapid refolding of a proline-rich all-beta-sheet fibronectin type III module. *Proc. Natl. Acad. Sci. USA* 93, 10703-10706.

6. Dickinson, C. D., Veerapandian, B., Dai, X.-P., et al. (1994) Crystal structure of the tenth type III cell adhesion module of human fibronectin. *J. Mol. Biol.* 236, 1079-1092.

7. Main, A, L., Harvey, T. S., Baron, M., Boyd, J., and Campbell, I. D. (1992) The three-dimensional struc-

ture of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* 71, 671-678.

8. Koide, A., Bailey, C. W., Huang, X., and Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol.* 284, 1141-1151.
9. Xu, L., Aha, P., Gu, K., et al. (2002) Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* 9, 933-942.
10. Koide, A., Abbatiello, S., Rothgery, L., and Koide, S. (2002) Probing protein conformational changes by using designer binding proteins: application to the estrogen receptor. *Proc. Natl. Acad. Sci. USA* 99, 1253-1258.
11. Sidhu, S. S., Lowman, H. B., Cunningham, B. C., and Wells, J. A. (2000) Phage display for selection of novel binding peptides. *Methods Enzymol.* 328, 333-363.
12. Origene Technologies, Inc. (1998) *Dup-LEXA Yeast Two-Hybrid System Manual.* Rockville, MD.
13. Golemis, E. and Serebriiskii, I. (1997) Two-hybrid system/interaction trap, in *Cells: A Laboratory Manual*, CSH Laboratory Press, Cold Spring Harbor, NY, pp. 69. 1-69. 40.
14. Chappel, J. A., He, M., and Kang, A. S. (1998) Modulation of antibody display on M13 filamentous phage. *J. Immunol. Methods* 221, 25-34.
15. Invitrogen. (2003) *Hybrid Hunter Manual.* Carlsbad, CA.
16. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and efficient sitedirected mutagenesis without phenotypic selection. *Methods Enzymol.* 154, 367-382.
17. Bio-Rad Laboratories. (1997) *Muta-Gene Mutagenesis Kit Manual.* Hercules, CA.
18. Gietz, R. D. and Woods, R. A. (2001) Genetic transformation of yeast. *Biotechniques* 30, 816-828.
19. Finley, R. L. Jr. and Brent, R. (1994) Interaction mating reveals binary and ternary connections between Drosophila cell cycle regulators. *Proc. Natl. Acad. Sci. USA* 91, 12980-12984.
20. Duttweiler, H. M. (1996) A highly sensitive and non-lethal beta-galactosidase plate assay for yeast. *Trends Genet.* 12, 340-341.

7　位点特异性核酸内切酶的蛋白质工程

Peter Friedhoff and Alfred Pingoud

概要　位点特异核酸内切酶涉及核酸生物化学的许多方面。限制酶及相关酶已成为作用于 DNA 的酶的典范。通过理性蛋白设计，投入了无数的精力试图改变其特异性。但是，从整体上说，成功很少，大概是由于其识别能力高度冗余，并且识别和催化紧密地耦合。本章描述少数几个成功改变特异性例子之一，即转换错配修复核酸酶 MutH，当被 MetS 和 MutL 刺激的时候，此酶在半甲基化位点 d（GATC）切开一缺口——转化为切割全甲基化以及半甲基化和未甲基化 DNA 的变体。本章将描述此设计项目中所涉及的各种步骤，从对结构的分析和对负责感知甲基化状态的候选氨基酸残基的鉴别，到产生和鉴定针对半甲基化 d（GATC）位点具有不同特异性的 MutH 变体。

关键词　限制性内切核酸酶；DNA 错配修复；DNA 甲基化；MutH；蛋白质工程；定点突变；荧光标记；寡核苷酸；毛细管电泳

7.1　介　　绍

类型Ⅱ限制性内切核酸酶是定点切割 DNA 不可或缺的工具。它们通常识别长度在 4～8 个 bp（碱基对）的双链 DNA 序列。在 Mg^{2+} 存在的条件下，它们在识别位点内或紧靠识别位点的地方切断 DNA 双链，以产生带有 5′-磷酸和 3′-羟基的平末端或黏末端（见参考文献［1］的述评）。限制性酶属于已知最具特异性的酶：对只差一个碱基对的序列的切割比对正常识别序列的切割要慢 10^6 倍。这种极端的精确性来自于直接阅读（与碱基相互作用）和间接阅读（与糖-磷酸骨架相互作用）的密切配合。典型地，在限制性酶和识别序列的碱基之间，除了无数与碱基接触形成的范德华力，和有些是靠水做中介的与骨架作用形成的氢键外，还形成了 15～20 个氢键。这意味着识别是多因素决定的和冗余的。因而，通过理性蛋白设计来改变限制性酶的特异性，由于这些原因而不容易成功。因为必须一次改变几个接触，并替换一个以上的氨基酸残基[2]。而且，在大多数限制性酶中，涉及识别的氨基酸残基位于很接近负责催化的残基位点。虽然这保证了识别与催化的紧密耦合，但一直是蛋白质工程的大障碍（见参考文献［3］的述评），因而回顾起来并不奇怪，理性蛋白设计并未做到真正的特异性改变[4]，无论是在试管中[5,6]，还是在体内[7,8]，研究者已转向用进化方法来改变限制性酶的特异性。但是即使使用这些方法，至今仅有特异性减少（如 *Bsty*I，从 RGATCY 到更优的 AGATCT，参考文献［7］），或特异性扩展（如 *Eco*57I，从 CTGAAG 到 CTGRAG，参考文献［8］）可以实现。这很可能是因为限制性酶在结构上没有足够的可变性，并且，为实现真正的特异性改变，需要多个氨基酸替换。这可能是一种保障，以防止不期望的、多数情况下是有害的、主基因组被限制性酶变体切除。而这些变体在翻译过程中

在小范围内总是被制造出来。通过同样的论证，可以预料，可能产生特性改变了的限制性酶变体，并且这些特性在进化中没有受到负选择。事实上，可以制造出含有简单氨基酸替换的 *Eco*RI 和 *Eco*RV 变体。这些变体可以比野生型酶更好地切除含有修饰碱基的 DNA 序列（*Eco*RI Q115，参考文献 [9]；*Eco*RV N188，参考文献 [10]）或骨架被修饰（*Eco*RV T94V，参考文献 [11]）的 DNA。

这里，我们报告一个与限制性酶相关的酶，即来自于大肠杆菌的错配修复蛋白 MutH 通过单氨基酸替换，怎样被转变为放宽了特异性的变体。使用过的操作步骤，在此得到详细描述，应该适合于广泛应用。首先，介绍了被研究的体系（见 7.1.1），同时还交代了结构上的考虑，以选择恰当的氨基酸替换，使特异性得到期望的变化（见 7.1.2）。

7.1.1　DNA 错配修复蛋白 MutH

在几乎所有有机体中，DNA 错配修复对纠正复制中的错误，都是最重要的过程之一；它使复制保真度提高了 1000 倍（见参考文献 [12] 的述评）。在大肠杆菌中，DNA 错配修复由 MutS 结合于错配位点引发，接着是 MutH 依赖于 MutL 的激活。激活的 MutH 在半甲基化的 d（GATC）位点切开未甲基化的（即刚复制的）链。这个位点可能在错配位点的上游或下游几百个碱基对处。切开的 DNA 链在螺旋酶Ⅱ的协助下，被四个核酸外切酶之一以外切方式降解，直到并超过错配位点。被移除的片段随后被 DNA 聚合酶Ⅲ和 DNA 连接酶合成新片段取代（和修复）。

MutH 是序列特异的 Mg^{2+} 依赖的 DNA 内切核酸酶，分子质量为 28kDa。它与类型Ⅱ限制性内切核酸酶 *Sau*3AI 有序列相似性[13]。*Sau*3AI 也识别 d（GATC）序列，并在 G 的 5′端切开。但是，MutH 只切割未被甲基化的链，无论是在半甲基化的或是在未甲基化的 DNA 中；而 *Sau*3AI 不管 DNA 是否处于甲基化状态两条链都切割。与典型的限制性内切核酸酶和 *Sau*3AI 都不同，MutH 作为单体是活性的。MutH 结构的确定（见参考文献 [14] 和 [15]）证实了 MutH 和类型Ⅱ限制性内切核酸酶 PD...D/EXK 家族间的关系。基于 MutH 和 *Pvu*II[16,17] 间的结构同源性，通过单体 *Mut*H 与齐二聚 *Pvu*II 的一个亚单位的叠合，鉴别出了 MutH 的假设的 DNA 结合和活性位点[13] 由 Asp70、Glu77 和 Lys79，以及其他残基组成[18~21]。

7.1.2 MutH 中涉及感知 DNA 的 d（GATC）位点甲基化状态的候选氨基酸残基的鉴别

我们使用了两个来源的信息以鉴别 MutH 中的那些氨基酸残基，它们作为可能的候选者，用来感知 d（GATC）位点的甲基化状态，即一条链上，甲基在腺嘌呤 N6 上存在，而在另一条链上则不存在这样的甲基。

已知在大肠杆菌、流感嗜血杆菌、霍乱弧菌、*Shewanella oneidensis* 和 *Colwellia* sp. 中都鉴定出了 MutH 蛋白；有可能在这些高度同源的蛋白质中，那些负责感知 d（GATC）位甲基化状态的氨基酸残基是保守的。MutH 与限制性酶，如来自于金黄色葡萄球菌的 *Sau*3AI 存在序列相似性。与 MutH 相似，这些酶识别 d（GATC）位点。与不切割甲基化 DNA（半甲基化的 DNA 只有未甲基化的链上被切开）的 MutH 不同，这些酶在未甲基化、半甲基化和全甲基化的 DNA 的两条链上都切割。MutH 酶和限制

性酶序列的逐点比较，应该使得两个家族的蛋白氨基酸残基的鉴别对序列识别同等重要，以及只应在 MutH 蛋白中存在感知甲基化状态所需要的残基。

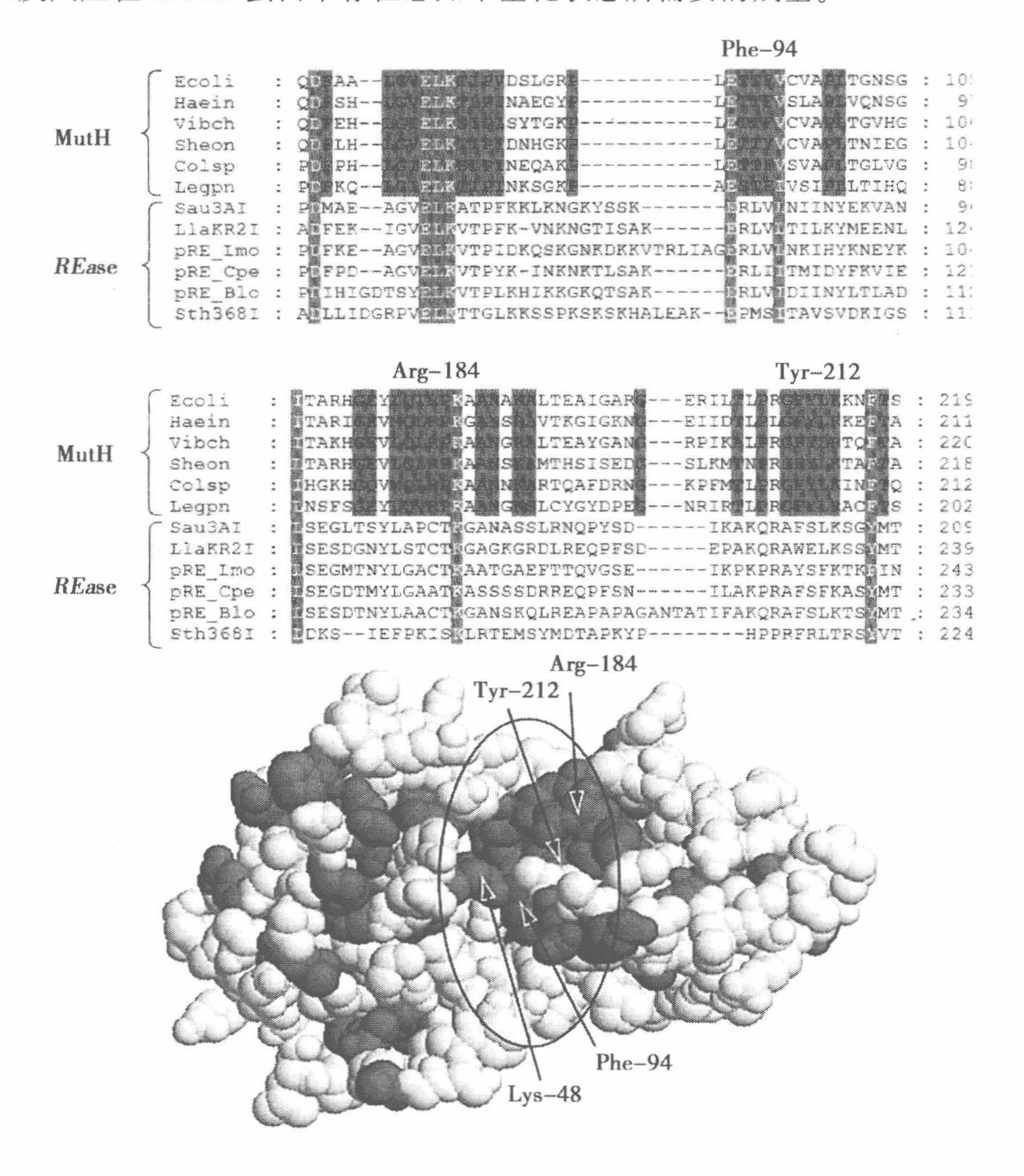


图 7.1 在 MutH 结构上组间特异性残基图示。上图：MutH 蛋白组与 *REase* 限制性内切核酸酶组的序列比对部分结果。两个组都保守的残基用灰底白字母标出。只在 MutH 组保守的残基用灰底黑字母标出。下图：在大肠杆菌 MutH 蛋白结构中，组间特异性残基用灰色示出。那些涉及直接接触甲基化碱基的残基位置及名称被进一步标示出来。

蛋白质氨基酸序列的比对证明这些核酸酶共享许多保守的氨基酸残基，由此可以推测这些保守残基都涉及共同的功能，包括 DNA 结合、识别和剪切。某些氨基酸残基只在 MutH 蛋白中保守，其中有 Phe94、Arg184 和 Tyr212（图 7.1）。为说明这些氨基酸残基中的哪些有可能位于蛋白质-DNA 界面上，用催化中心的残基作为参考点，我们把 MutH 的结构叠加到限制性酶-DNA 复合物结构之上。在重叠的结构中，下列残基接近对应于 DNA 序列双链中的腺嘌呤残基的核酸碱基：Lys48 面对小沟，Phe94、Arg184 和 Tyr212 面对大沟（图 7.2）。因为腺嘌呤残基的 N6 位于大沟，只有 Phe94、Arg184 和 Tyr212 是好的候选残基，它们可用于感知一条链中的 N6 的甲基化和另一条链 N6 的未甲基化。


Tyr212 似乎特别地引人注目，因为叠加结果显示其位置接近两条链中的腺嘌呤残基。

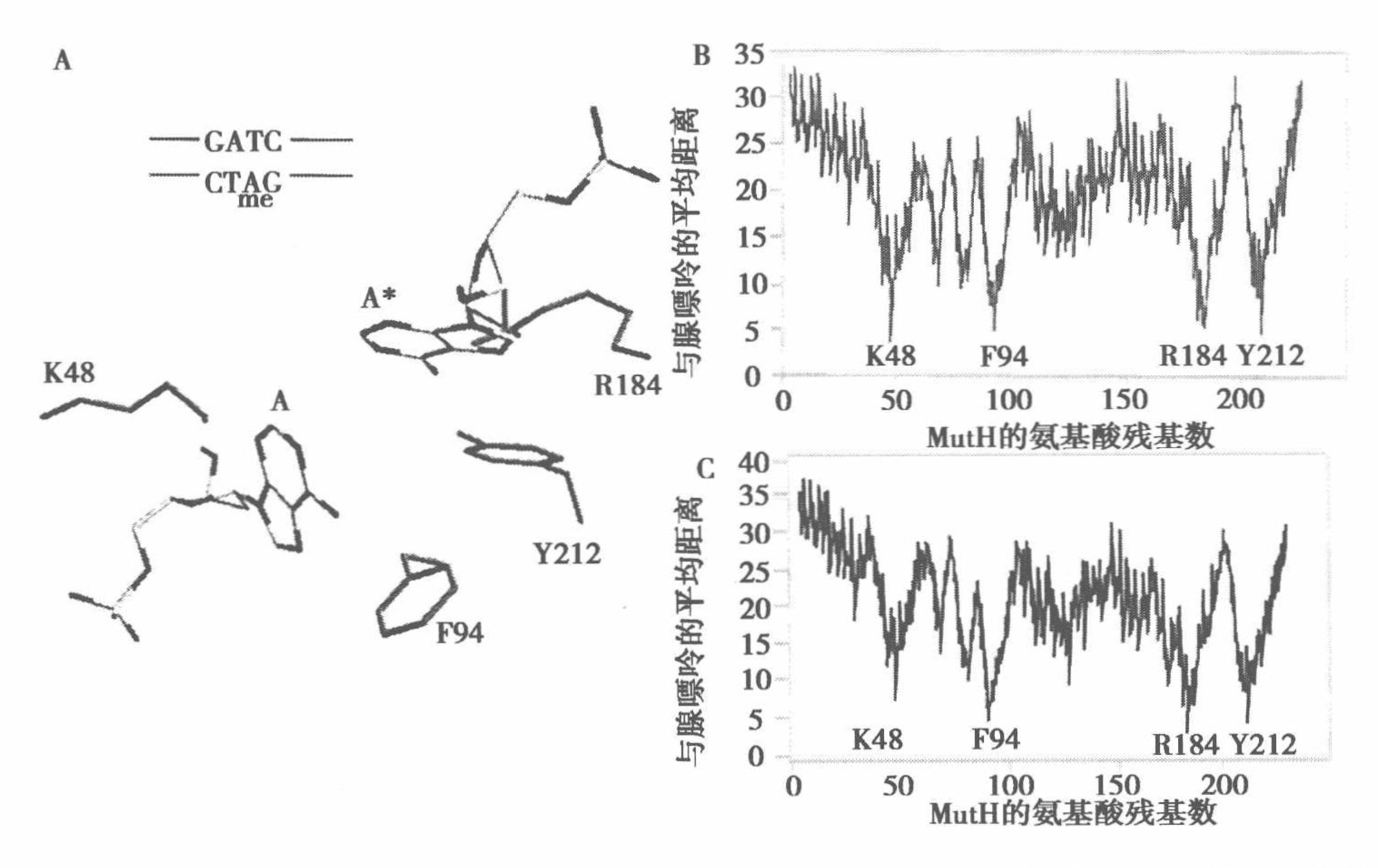


图 7.2　MutH 与限制性内切核酸酶的结构叠加。A. MunI 识别序列中相应 MutH 识别序列 d（GATC）两个腺嘌呤的两个核苷酸如图所示。标注有星号（A*）的核苷酸相应于 MutH 识别序列中下方核酸链的腺嘌呤，它可以被甲基化并且介导上方核酸链的剪切。B，C. MutH 与 11 种不同限制性内切核酸酶（*Bam*HI、*Bg*lⅠ、*Bg*lⅡ、*Bso*BI、*Eco*RI、*Eco*RV、*Hinc*II、*Mun*I、*Nae*I、*NgoM*IV 和 *Pvu*II）结构叠加的定量分析。图示分别为 MutH 结构中原子与相应于 MutH 识别序列上方链腺苷酸中　B. 或者下方链腺苷酸中　C. 的 N6（腺嘌呤）、O6（鸟嘌呤）、N4（胞嘧啶）或者 O4（胸腺嘧啶）原子的平均距离。注意图中的极小值指示了 MutH 中氨基酸残基如 K48、F94、R184 和 Y212 与所有叠加的 DNA 结构中相应于上、下方核酸链中腺嘌呤 N6 的接近程度。

7.1.3　MutH 的 DNA 切口和切割活性分析

MutH 和 MutH 变体在 DNA 切口和切割中的活性和特异性的分析，最好用含有单个未甲基化、半甲基化和全甲基化 d（GATC）位点的双链 DNA 作为底物来研究。为探测两条链中磷酯键的切割，DNA 底物的两条链应带有不同的标记。这样的底物可以用荧光标记引物进行聚合酶链反应（PCR）来制备。在 d（GATC）位点甲基的引入需要 dam 甲基化酶做酶性甲基化。通过 λ 外切酶降解全甲基化底物的两条甲基化链中的一条，并与一条互补的未被甲基化的链杂交，就得到半甲基化的底物。对这些底物与 MutH 保育后的切割产物的分析，可以很容易地用带有激光诱导荧光探测的变性聚丙烯酰胺毛细管电泳来实行[19,22]。

7.2　材　　料

7.2.1　DNA 结合和识别残基的鉴别

以下计算程序是用来鉴别 MutH 涉及感知 d（GATC）位点甲基化状态的残基。

（1）ClustaW（http：//www.ebi.ac.uk/clustalw/）。

（2）ClustalX（http：//bips. u-strasbg. fr/fr/Documentation/ClustalX/；请注意，网址有大小写区别）。

（3）GeneDoc（http：//www. psc. edu/biomed/genedoc/）。

（4）RasMol（http：//www. openrasmol. org/）。

（5）Swiss PDB Viewer（http：//www. expasy. org/spdbv/）。

7.2.2 MutH 的定点突变

（1）pMQ402——BAD18 的一种衍生物，用作细菌表达质粒[23]。在诱导蛋白质表达的阿拉伯糖启动子的控制之下，它含有用 His 标记的 MutH 蛋白的编码序列。表达出的蛋白质可用 Ni-次氮基三乙酸（NTA）琼脂糖亲和柱纯化。质粒含有供筛选的氨苄青霉素抗性基因。

（2）PCR 引物，用于定点突变。

（3）*pfu* 聚合酶。

（4）三磷酸脱氧核糖核苷。

7.2.3 过表达、纯化和鉴定

（1）Luria-Bertani（LB）培养基（每升）：10g 菌用胰蛋白胨，5g 菌用酵母抽提物和 5g 氯化钠。用氢氧化钠调节 pH 至 7.2～7.5，并灭菌。

（2）阿拉伯糖（Sigma）。

（3）苯甲基磺酰氟（PMSF）（Merck）。

（4）Ni-NTA 琼脂糖（Qiagen）。

（5）大肠杆菌菌株 XL-1 Blue MRF'（Stratagene）。

（6）超声破碎器。

（7）冷冻离心机。

（8）空柱（BioRad）。

（9）结合缓冲液：20mmol/L Tris-氯化氢、5mmol/L 咪唑、1mol/L 氯化钠和 1mmol/L PMSF，pH7.9。

（10）清洗缓冲液：20mmol/L Tris-氯化氢、60mmol/L 咪唑、1mol/L 氯化钠和 1mmol/L PMSF，pH7.9。

（11）洗脱缓冲液：20mmol/L Tris-氯化氢、200mmol/L 咪唑、1mol/L 氯化钠和 1mmol/L PMSF，pH7.9。

（12）透析缓冲液：10mmol/L HEPES-氢氧化钾、100mmol/L 氯化钾、1mmol/L EDTA 和 1mmol/L 二硫苏糖醇，pH7.9。

（13）透析缓冲液 G：10mmol/L HEPES-氢氧化钾、100mmol/L 氯化钾、1mmol/L EDTA、1mmol/L 二硫苏糖醇和 50%甘油，pH7.9。

7.2.4 内切酶活性测试

（1）测试缓冲液：10mmol/L Tris-氯化氢，pH7.5、10mmol/L 氯化镁、0.75mmol/L 三磷酸腺苷和 0.1mg/ml 牛血清白蛋白。

（2）大肠杆菌 MutL 蛋白。

(3) 荧光团标记的寡核苷酸或荧光标记的 PCR 产物［模板 DNA、磷酸化引物、荧光团标记引物、*dam* 甲基化酶（NEB）、λ-外切酶（NEB）、外切酶 I（NEB）、PCR 自旋柱（Qiagen）和 *pfu* DNA 聚合酶（Promega)］。

(4) 模板抑制剂（Perkin-Elmer）。

(5) GeneScan-500 TAMRA DNA 分子质量标准（Perkin-Elmer）。

(6) ABI PRISM 310 基因分析仪（Perkin-Elmer）。

(7) 灌有加了 8mol/L 尿素的 POP-4 聚合物的 47-cm 毛细管（内径：50μm）（Perkin-Elmer）。

(8) 加有 1mmol /L EDTA 的 1×基因分析缓冲液（Perkin-Elmer）。

7.3 方 法

此处列出的方法概括了为定点突变而做的氨基酸位点选择（见 7.3.1）、为在细菌细胞中表达而做的 MutH 蛋白突变（见 7.3.2）和 MutH 变体的鉴定（见 7.3.3）。

7.3.1 为定点突变对氨基酸的选择

从生物技术信息国家中心基因组基本局部比对搜索工具（basic local alignment search tool，BLAST）页面（http：//www.ncbi.nlm.nih.gov/BLAST/）[24]得到 MutH 蛋白和相关限制性内切核酸酶的序列。更多的序列从 PSI BLAST 服务器[25]得到。使用 PAM250 矩阵，用 ClustalX 程序比对序列[26]。比对好的序列用 GeneDoc 程序分析[27]。程序的使用方法可在程序内使用“Help”功能得到。结构的坐标从 PDB 数据库（http：//www.rcsb.org/pdb/）得到。程序 RasWin 和 Swiss PDB viewer 用于结构观察。

7.3.1.1 与 DNA 结合有关的保守残基团的辨别

在序列比对（如用 ClustalW 或 ClustalX）之后，完成下列步骤。

(1) 将比对结果输入 GeneDoc 程序（见 7.2.1；参考文献［27］）。

(2) 用“Group”功能将序列分组（如一个组包含 MutH，而另一个组包含限制性内切核酸酶）。

(3) 鉴别组特异的残基（见注 1）。

(4) 用 GeneDoc 内的“RasMol Script Dialog”功能，生成 RasMol 执行文稿，以将组特异残基映射入（标记在）MutH 蛋白结构中。

(5) 将 MutH 结构载入 RasMol 程序（见 7.2.1；参考文献［28］)，在“RasMol Command Line”用“Script”命令执行从 GeneDoc 输出的文稿，以观察组特异残基。

(6) 辨别位于假设为 DNA 结合位点的候选残基。

7.3.1.2 辨别接触特定碱基的残基

(1) 下载限制性内切核酸酶与 DNA 底物（或产物）形成复合物的现有序列。

(2) 运行 Swiss PDB viewer，以 3 个催化残基（如 MutH 的 D70、E77 和 K79）的主链原子为种子，将限制性内切核酸酶结构与目标 MutH 结构进行匹配。

(3) 用“Improve fit”功能增强拟合效果。

（4）将重叠结构的坐标输出到电子制表程序。

（5）计算目标蛋白（如 MutH）的任意原子到重叠结构中碱基的距离，此碱基对应于目标蛋白中的目标碱基。

（6）对所有重叠结构，重复这些步骤。

（7）计算平均距离，并辨认对感兴趣碱基距离最近的残基。

（8）将结果与组特异残基分析比较。

（9）为定点突变选择有希望的残基。

7.3.2 MutH 变体定点突变

MutH 变体的克隆，采用如 Kirsch 和 Joly 所述[29]修改过的 QuikChange 操作规程（Stratagen），用质粒 pMQ402（来自 Dr. M. G.，University of Massachusetts Medical School，MA）为模板，以及两个寡聚脱氧核苷酸用于突变以适合于产生长度在 50～500bp 的 PCR 产物（见注 2）。

7.3.3 MutH 变体的纯化和鉴定

为了适合于体外测试，MutH 变体必须得到纯化。

7.3.3.1 从细菌细胞纯化 MutH 变体

（1）用细菌质粒以标准分子生物学方法转化 XL-1 Blue MRF 细胞。

（2）将细胞平铺在含有氨苄青霉素的 LB 培养板上，于 37℃过夜培养。

（3）选一单菌落，在 500ml 含 75μg/ml 氨苄青霉素的 LB 培养液中，于 37℃生长。

（4）在 600nm 处光密度值为 0.8 时，于 28℃，用 0.2%（*m/V*）阿拉伯糖（终浓度）诱导 2.5h（见注 3）。

（5）以 3000*g* 离心细胞 10min。

（6）重悬细胞团于 10ml 结合缓冲液中［见 7.2.3（9)］。

（7）用 Branson 超声机，输出级 5，占空率 50%，超声破碎悬浮物 5 次，每次 1min。每次超声间歇冷却溶液 1min。

（8）在冷冻离心机上，以 30 000*g* 离心细胞碎片 30min。

（9）于 4℃，将上清液与 1.5ml Ni-NTA 琼脂糖浆柔和地混合 30min。

（10）将 Ni-NTA 琼脂糖移至一空柱中。

（11）用 20ml 清洗缓冲液冲洗该柱［见 7.2.3（10)］。

（12）用 0.5ml 洗脱缓冲液［见 7.2.3（11)］洗脱。没有必要切除 His 标记，因其不干扰内切酶活性（见注 4）。

（13）以在 280nm 处的光密度值作为判断依据，收集合并得到的每份蛋白质。

（14）于 4℃，用 500ml 透析缓冲液［见 7.2.3（12)］透析样品最少 2h。换缓冲液 2 次。

（15）用 500ml 透析缓冲液 G［见 7.2.3（13)］透析样品。

（16）在透析缓冲液［见 7.2.3（12)］中以 1∶10 比例稀释样品，测量 280nm 处的光吸收，以便用理论消光系数计算 MutH 的摩尔浓度[30]。

（17）于－20℃保存蛋白质。

7.3.3.2　MutH 的切割分析

用含有未甲基化、半甲基化或全甲基化的单个 d（GATC）位点（图 7.3）的 DNA 底物（用它处所描述的方法合成的寡核苷酸，或者 PCR 产物，参考文献［19］和［20］）来开展 MutH 和 MutH 变体的切割分析。用不同的荧光染料（分别用 FAM 和 TET）标记底物上下两端的链，以便检测每条链上的切割。

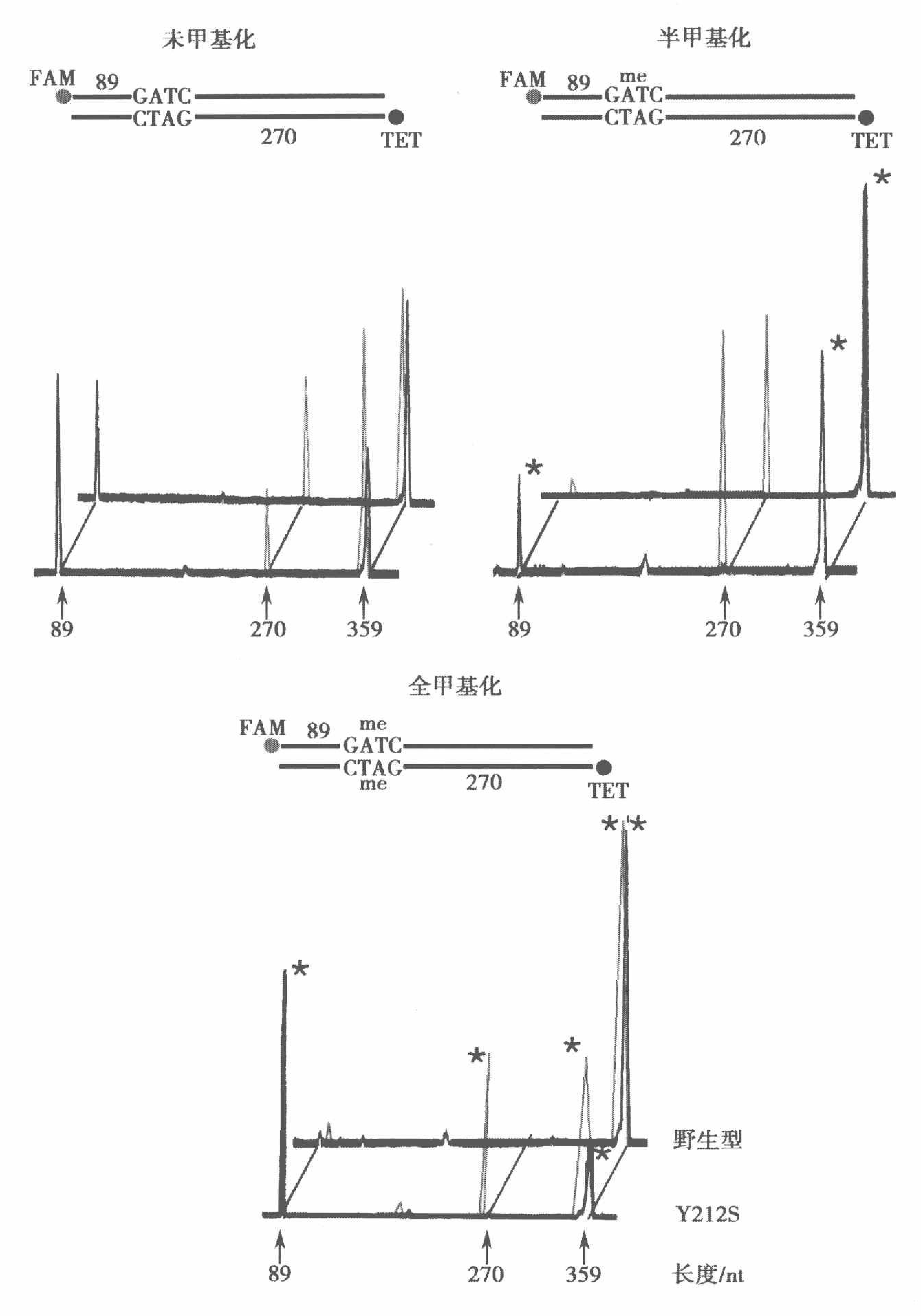


图 7.3　利用未甲基化、半甲基化和全甲基化 DNA 底物进行 MutH 野生型与突变型 Y212S 的剪切分析。在上方链标记了 FAM（灰色）和下方链 TET（黑色）的 359bp 长且含有单个 d（GATC）位点（未甲基化、半甲基化或者全甲基化）底物与野生型的 MutH 与突变型 Y212S 与 MutL 一起孵育。剪切反应利用毛细管电泳分析。底物（359）与产物（89 及 270）链的位置由箭头标示出。甲基化 DNA 链由星号标出。

(1) 10nmol/L 浓度的 DNA 底物在 10μl 含有 500nmol/L 的 MutL 和浓度在 10～500nmol/L 的 MutH 的测试缓冲液（见注 1；误，似应为 7.2.4 节，第 1 条——译者）中保温（见注 5）。

(2) 反应混合物于 37℃温育。

(3) 以适当的时间间隔（10s 到 30min）分装反应混合物，每份含 25fmol PCR 产物，与 12μl 模板抑制剂（Perkin-Elmer）和 0.5μl Gene-Scan-500 TAMRA size standard（Perkin-Elmer）充分混合。

(4) 加热到 95℃2min，并立即在冰上冷却。

(5) 在配有 47cm（内径 50μm）毛细管（其中含有加了 8mol/L 尿素的 POP-4 聚合物（Perkin-Elmer））的 ABI PRISM 310 基因分析仪（Perkin-Elmer）上分析样品。

(6) 用 5s，将样品电注入毛细管，使用电压 15 000V，并在 15 000V 和 60℃、使用 1×基因分析缓冲液外加 1mmol/L EDTA（Oerkin-Elmer）作为电极缓冲液，使用 30min 完成跑电泳过程。

(7) 记录切割的和未切割的荧光标记的 DNA 数量（图 7.3）。

(8) 测定切割速度。

7.4 注

1. 通过在“Edit Sequence Groups”对话框中设置“Group Cons Level”和“PCR Maximum Level”可以改变图像阴影。

2. 一个引物会导致期望的突变，而另一个为反义引物，以生成所谓“megaprimer”用于定点突变（据 QuikChange 操作规程）。

3. 诱导的时间和温度取决于被研究的体系。

4. 某些情况下，需切除 His 标记。此处使用的建构允许在 His 标记和 N 端甲硫氨酸间的凝血酶位点做切割。

5. 与用 MutL 活化不同的另一种方法是，加入 10%（*V*/*V*）二甲基亚砜，可使 MutH 内切酶活性激活到 10 倍。

致谢 作者实验室内的工作得到 Herbert Stolzenberg Stiftung，Deutsche Forschungsgemeinschaft（Pi 122/12-4），和 Fonds der Chemischen Industrie 的支持。

参考文献

1. Pingoud, A. and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* 29, 3705-3727.

2. Lanio, T., Jeltsch, A., and Pingoud, A. (2000) On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing EcoRV as the target. *Protein Eng.* 13, 275-281.

3. Jeltsch, A., Wenz, C., Wende, W., Selent, U., and Pingoud, A. (1996) Engineering novel restriction endonucleases—principles and applications. *Trends Biotech.* 14, 235-238.

4. Alves, J. and Vennekohl, P. (2004) in *Restriction Endonucleases* (Pingoud, A., ed.), Springer, Berlin.

5. Lanio, T., Jeltsch, A., and Pingoud, A. (1998) Towards the design of rare cutting restriction endonucleases: using directed evolution to generate variants of *Eco*RV differing in their substrate specificity by two orders of magnitude. *J. Mol. Biol.* 283, 59-69.

6. Lanio, T., Jeltsch, A., and Pingoud, A. (2002) in *Directed Molecular Evolution of Proteins* (Brakmann, S. and Johnsson, K., eds.), Wiley-VCH, Weinheim, Germany, pp. 309-327.
7. Samuelson, J. C. and Xu, S. Y. (2002) Directed evolution of restriction endonuclease BstYI to achieve increased substrate specificity *J. Mol. Biol.* 319, 673-683.
8. Rimseliene, R., Maneliene, Z., Lubys, A., and Janulaitis, A. (2003) Engineering of restriction endonucleases: using methylation activity of the bifunctional endonuclease Eco57I to select the mutant with a novel sequence specificity. *J. Mol. Biol.* 327, 383-391.
9. Jeltsch, A., Alves, J., Oelgeschlager, T., Wolfes, H., Maass, G., and Pingoud, A. (1993) Mutational analysis of the function of Gln115 in the EcoRI restriction endonuclease, a critical amino acid for recognition of the inner thymidine residue in the sequence -GAATTC- and for coupling specific DNA binding to catalysis. *J. Mol. Biol.* 229, 221-234.
10. Wenz, C., Selent, U., Wende, W., Jeltsch, A., Wolfes, H., and Pingoud, A. (1994) Protein engineering of the restriction endonuclease *Eco*RV: replacement of an amino acid residue in the DNA binding site leads to an altered selectivity towards unmodified and modified substrates. *Biochim. Biophys. Acta* 1219, 73-80.
11. Lanio, T., Selent, U., Wenz, C., et al. (1996) EcoRV-T94C—a mutant restriction endonuclease with an altered substrate specificity towards modified oligodeoxynucleotides. *Protein Eng.* 9, 1005-1010.
12. Modrich, P. and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* 65, 101-133.
13. Ban, C. and Yang, W. (1998) Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.* 17, 1526-1534.
14. Yang, W. (2000) Structure and function of mismatch repair proteins. *Mutat. Res.* 460, 245-256.
15. Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y., and Morikawa, K. (2001) Crystal structure of the archaeal holliday junction resolvase Hjc and implications for DNA recognition. *Structure* 9, 197-204.
16. Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P. A., Wilson, K. S., and Kokkinidis, M. (1994) Crystal structure of *Pvu*II endonuclease reveals extensive structural homologies to EcoRV. *Nat. Struct. Biol.* 1, 469-475.
17. Cheng, X., Balendiran, K., Schildkraut, I., and Anderson, J. E. (1994) Structure of PvuII endonuclease with cognate DNA. *EMBO J.* 13, 3927-3935.
18. Loh, T., Murphy, K. C., and Marinus, M. G. (2001) Mutational analysis of the MutH protein from *Escherichia coli*. *J. Biol. Chem.* 276, 12, 113-12, 119.
19. Friedhoff, P., Thomas, E., and Pingoud, A. (2003) Tyr-212: a key residue involved in strand discrimination by the DNA mismatch repair endonuclease MutH. *J. Mol. Biol.* 325, 285-297.
20. Friedhoff, P., Sheybani, B., Thomas, E., Merz, C., and Pingoud, A. (2002) *Haemophilus influenzae* and *Vibrio cholerae* genes for *mutH* are able to fully complement a *mutH* defect in *Escherichia coli*. *FEMS Microbiol. Lett.* 208, 121-126.
21. Junop, M. S., Yang, W., Funchain, P., Clendenin, W., and Miller, J. H. (2003) In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. *DNA Repair (Amst.)* 2, 387-405.
22. Thomas, E., Pingoud, A., and Friedhoff, P. (2002) An efficient method for the preparation of long heteroduplex DNA as substrate for mismatch repair by the *Escherichia coli* MutHLS system. *Biol. Chem.* 383, 1459-1462.
23. Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* 177, 4121-4130.
24. Wheeler, D. L., Church, D. M., Lash, A. E., et al. (2002) Database resources of the National Center for Biotechnology Information: 2002 update. *Nucleic Acids Res.* 30, 13-16.
25. Altschul, S. F., Madden, T. L., Schaffer, A. A., et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
26. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24,

4876-4882.

27. Nicholas, K. B., Nicholas, H. B. J., and Deerfield, D. W. I. (1997) GeneDoc: analysis and visualization of genetic variation. *EMBnet NEWS* 4, 14.

28. Sayle, R. A. and Milner-White, E. J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* 20, 374.

29. Kirsch, R. D. and Joly, E. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.* 26, 1848-1850.

30. Pace, C. N., Vajdos, F. Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 4, 2411-2423.

第二部分　蛋白质工程中的进化策略

8 蛋白质库的设计和筛选——概率计算

Michel Denault and Joelle N. Pelletier

概要 在设计蛋白质库进行筛选时，在能够进行筛选的物理极限内我们必须尽一切可能产生出蛋白质变体的多样性。本章的目标是将概率的语言引入蛋白质工程实验中，进而来回答一些常见的问题。例如，我们怎样才能最有效地设计蛋白质库？我们最终从一个多样性完备的理论蛋白质库（简称理论库）中取样的比率是多少？错过理论库中一个个体的概率是多少？我们选择的变体中的突变是否具有统计意义，或仅仅是随机变异的产物？在整个蛋白质工程实验过程中这些计算标准可以使我们更好地设计和评价我们得出的蛋白质变体库。

关键词 蛋白库筛选；实验偏差；进化库的表现度；泊松分布；假设的 χ^2 检验，多样性

8.1 简 介

自然进化是一个产生给定样品的很多“变体”随后在“变体”中选择那些最适于特定环境或特定目的的过程。自然进化是机体内遗传突变或重组进而产生新蛋白质变体的结果。抗体的产生很好的描述了这种进化过程：天然抗体库在机体内产生大量的各种“变体”，这些变体被不断地取样测试，那些最适于与特定配体结合的变体被从天然库中筛选出来进行下一轮的优化。蛋白质变体的选择需要基因型与表型的耦合——这在自然进化中是自然发生的，但在实验室进行人工蛋白质进化时却很难实现。在过去的 10 年中，随着很多策略的发展可以产生更好说明的大蛋白质库来联结基因型与表现型，与此同时也发展了许多新型功能筛选方法和选择策略，蛋白质工程的工具发生了革命性进展（见参考文献 [1] 和 [2]）。

本章的目标是将与蛋白质库的产生以及利用这些库进行筛选的策略相关的统计与概率计算以组合生物学家的语言进行阐述，以利于这些计算方法的应用。本章的目的也是要揭示这些计算与库的产生及筛选过程的相关性。当我们设计一个基于蛋白质库筛选的蛋白质工程实验时，必须平衡好产生一个最具多样性的蛋白质库的科学价值与实际筛选或选择策略方法局限性之间的关系。与蛋白质库表征相关的统计与概率问题是重要的，并且这些问题随着不同的应用而变化。我们还给出了实验偏差的检测，来帮助评价筛选前后蛋白质库的质量及偏差的出现情况。我们将会看到对于筛选库问题的简单分析可以揭示出很重要，甚至有时是违反直觉的结果，这些信息反过来可以使我们的实验设计得到快速改进，并且提供给我们解释结果的更强有力的依据。

我们首先讨论对于设计蛋白质库相关的一些重要参数，如蛋白质库的大小、组分的偏好性、密码子简并性及编码多样性。然后，我们将继续讨论一系列常见的问题，对于每个问题都给出相关的明晰的数学处理。为了阐明公式的应用我们通常会提供数值解范例。如果所需数学工具比较复杂费时，我们会给实验者提供 Excel 文档文件来输入合适的变量。我们所提出的问题都具有足够的普遍性适用于基于蛋白库的多样性系统而不只适用于一些特例。

我们必须提起注意下面给出的很多结果只适用于参数是足够大或足够小的情形，如果应用于其他地方则可能导致错误结论。我们也会提供一些特殊条件的情形。

基于蛋白质库策略数学处理的补充性工作还包括 DNA 混编（shuffling）的数学模型，特别是 Moore 和 Maranas（见参考文献［3］～［6］）及 Sun（见参考文献［7］）的重要工作，最近还有 Blackburn 及其同事[8]的工作。上述后者的工作也提供了库的等概率结果的库表征的数学处理；作为处理库的非等概率结果的库表征的开端，我们也提出了一个相同处理方法。参考文献［8］还提出了一个易错（error-prone）PCR 产生库的多样性的评估方法，更重要的是为用户提供了简捷的计算机程序。

8.1.1　库的产生及筛选的参数

蛋白库的设计及筛选过程中有一些参数必须事先确定，如所需库的大小，特定应用的库的表征以及筛选策略所附加的限制条件。尽管可能导致概念及实验误差，这些参数还是可以凭直觉提出。这些参数的数学意义将在 8.3.1 节中进一步讨论，以帮助实验者更好地计划及实施基于库的实验工作。

8.1.1.1　蛋白质库大小的设计

无论组合生物学家利用自然产生的蛋白质库还是合成库，首先必须确定库的大小，以利于将来能够精确评价筛选结果。自然产生的蛋白质库包括天然抗体集合以及来自某种细胞器（或器官）和某一机体的整个蛋白质组。合成库一般源于某一种蛋白质或一类同源蛋白，这些蛋白质可以是高度多样性的。当前用来产生这些合成库的技术包括饱和突变（saturation metagenesis）；利用 PCR 的随机突变；“基因混编”，即将类似来源的 DNA 片段组合在一起；“定向”突变，即改变特定区域的 DNA 编码；以及序列的非同源、随机重组策略等（非同源重组可参见文献［1］、［2］及［9］）。对于设计合意的合成库的大小，下列考虑是必需的。

对于小的多肽库（如 10 个氨基酸长的十肽），有 20^{10}（$\approx 1.02\times10^{13}$，原文错误——译者注）种可能的氨基酸组合。同样的计算适用于蛋白质中特定位置包含有 10 个随机分布的氨基酸的情形。然而，由于密码简并性，对于 10 个氨基酸的多肽存在着更多的编码可能性，我们必须产生大于 20^{10} 的 DNA 多样性库来编码一个合适的包括 20^{10} 种不同多肽的组合，这样才能根据每个多肽的编码可能性表征出其合理的分布。例如，在一个随机 DNA 库中，对于丝氨酸、精氨酸和亮氨酸（Ser、Arg 和 Leu，以下同，用英文三字母缩写来表示 20 种不同氨基酸）存在 6 种编码可能性，这是最高的冗余度。这样，编码一个 10-Ser 多肽是高度冗余的，有 $6^{10}=6\times10^{7}$ 可能的组合都编码这种 10-Ser 肽，然而对于非简并的 10-Met 多肽只存在唯一的编码组

合。(像 Met 这样) 非简并氨基酸只用 1/64 的三联码来唯一编码，密码简并性使得实际应用的 DNA 库必须大于不考虑简并性计算出来的结果以确保非简并氨基酸有一定的机会出现。更明确的说，在这个例子中我们必须编码 64^{10} ($\approx 1.2\times 10^{18}$) 种可能的 DNA (三联码) 序列来确保 10-Met 多肽可能有一定的概率出现。不幸的是，编码 1.2×10^{18} 种十肽的 DNA 库质量大约是 2mg，这样的量对于大多数研究机构来说是太大了。值得提出的是利用上述对于十肽编码的分析我们可以考虑更长的多肽或蛋白质其中任意 10 个氨基酸 (连续的或间隔的) 需要被改变的情形。因此，在计划构建多肽或蛋白质库时密码子的简并性是必须考虑到的。

8.1.1.2 偏好性蛋白质库

我们常常需要使 DNA 库具有编码偏好性。这样做最显然的优势是：①降低终止密码子的影响，终止子在非偏好性库中会产生大量非功能性序列 (表 8.1)；②使 (DNA) 库向需要的氨基酸组分倾向；③降低不同氨基酸在密码子表征上的差别，使得各种氨基酸在库中具有更加均一的表征。具有 “NNC” 或 “NNT” (N 表示 4 种核苷酸中任意一种) 重复序列的 DNA 库可以使密码子从 64 种变为 16 种，同时消除了 3 种终止子 (表 8.1)。然而这样做使得 5 种氨基酸 (Met、Trp、Gln、Glu 和 Lys) 没有被编码，导致了编码多样性的损失。利用 “NNC” 或 “NNT” 重复编码的缺点是某些密码子在大肠杆菌 (最常用蛋白质表达宿主) 中密码子应用偏好性较差[10]。“NNC/T” 编码也是常用的，它补偿了密码子利用偏好性问题，但使密码子数目增加了一倍而没有增加编码的氨基酸种类。因为在完全随机的编码库中 (“NNN”)，得到没有终止子产物的概率随着简并位点的增加而急剧下降 (表 8.1) 从而导致低质量库的产生，所以上述无终止子的编码是完全值得牺牲一些氨基酸多样性的。另外的做法是用 “NNC/G” 或 “NNT/G”，这样所有 20 种氨基酸都被编码且只包含一个终止子；“NNC/G” 密码子对大肠杆菌具有较好的应用性。这样，得到无终止子产物的概率显著大于用 “NNN” 编码的库。因此，当需要产生多重简并性氨基酸时，序列中的终止子对于库的质量是非常有害的，如果所用的筛选策略是需时费力并且很昂贵的话，应该尽力避免终止子。

表 8.1

偏好密码子编码	密码子	氨基酸	终止子	无终止子插入的概率	
				在含有 10 个偏好密码子的蛋白质中	在含有 30 个偏好密码子的蛋白质中
NNN	64	20	3	0.62[a,b]	0.24
NNC or NNT	16	15	0	1	1
NNC/G or NNT/G	32	20	1	0.73	0.39

a. 数值 0.62 表示对含有 10 个随机密码子 (NNN) 的蛋白质或多肽来说，预计有 62% 的变量将是全长不含有终止子的。

b. 由 $(61/64)^{10}=0.619$ 计算得到，其中对 10 个密码子来说，有 61/64 的密码子是不编码终止子的。

实验者经常需要在特定的位置产生特殊的氨基酸序列。尽管有现成软件来实现这一工作，我们还是可以手工完成这样的寡核苷酸偏好性设计。例如，T. J. Magliery（www. chemistry. ohio-state. edu/～magliery/publicatioin. html）写的“Mixed Codon Worksheet”就是一个基于 Excel 文件的设计程序，把一个需要的氨基酸列入表中，不同的可能性会被计算出来，使得实验者可以根据需要选择最合适的密码子。一些寡核苷酸合成供应商也可以根据需要提供各种可能比例的核苷酸，使得实验者可以偏好性选择核苷酸分布从而使特定氨基酸得到偏好。这种策略可以按照需要用来减少或者增加编码表征中的差异性。

在理想状况下，一个具有偏好性的库可以从“三核苷酸联码”中构建[11]，其中 3 个特定核苷酸共价连接在一起构成“砌块”（building block），每个“砌块”编码一个特定氨基酸，并且可以被用来以需要的比例混合放在特定的位置进行寡核苷酸合成。由于这种做法是非简并的，所以可以将 DNA 库的大小减至正好和感兴趣的蛋白质库大小一样。利用三核苷酸的优势已经被特别提出（见参考文献［12］～［14］），但由于这种方法近来只是处于商业化过程中（Glen Research Sterling，VA；及 Metkinen Oy，Kuusisto，Finland），完成的工作还相对较少。

8.1.1.3　库大小的限制

在具体实践中，可实现的库大小上限为毫克量级，这对于 30bp（相应于 10 个氨基酸）的双链 DNA 分子来说大约为 3×10^{16} 个。然而，蛋白质表达一般是基于载体 DNA 来实现的。一个典型的表达载体大概是 3000bp 长。因此，在毫克量级的理想状况下，对于一个 3000bp 的载体可以最多产生约 3×10^{14} 个双链 DNA 分子。这样，现实中对于 10 个完全随机氨基酸位置（1.2×10^{18} 种可能性）的上限仅仅是所有可能性的一小部分（更多关于库表征计算的细节参见 8.3.1.2）。这样，我们只能够部分的探索完全随机的十肽库（或者编码 10 个完全随机的氨基酸的蛋白质），完全不可能全面研究一个哪怕仅仅是平均大小的蛋白质。这意味着需要用两种方式产生蛋白质库。第一种，靶蛋白本身可被“完全”随机化；这种情形下，所得蛋白质库仅代表所有可能性的一个子集（见 8.1.2）。如果我们需要了解一个给定功能的野生型蛋白质序列中的相对重要性信息的话，这种做法会非常有效。筛选出的蛋白质中那些统计学上不常变化的残基提示其对于保留所选择的性质具有关键意义；这种提示很像利用多重天然蛋白序列比对来确定功能氨基酸残基。在第二种方法中，序列随机性被加以限制；仅仅特定残基（无论是否邻近）被完全随机化和（或）偏好性被引入随机化中以限制可能性数目。这样可以得到性质和大小完全确定的蛋白质库；而且库的大小可以被 8.1.1 节所述方法限制。如果我们需要对于库中的每个成员进行筛选（而不仅仅从一个子集合中取样）的话，则库的大小还可以进一步被限制。

8.1.2　蛋白质库的表征

与所需库的大小（内容）相关联的是我们要考虑对于特定应用的库表征。要使一个库的筛选效率高，编码的独立变体数必须与筛选策略及能力相匹配。如 8.1.1 节所述，产生大量和多样性的 DNA 库本身并不难。实际上我们必须限制库的大小才

能使之能够被充分地筛选。筛选所需表型（与基因型相关联）的工具相对于产生DNA库更加难以发展，这是因为这些工具不像操纵DNA的方法那样可以广泛应用。因此，筛选具有不同功能的蛋白质，必须建立起多样化的筛选策略（见参见文献[2]）。

一般来说，实验者进行手工操作可以筛选成百上千种蛋白质变体；自动化可以增加几个数量级的蛋白质变体个数。体内方法可以快速筛选几千到几十亿个蛋白质变体。无细胞（cell-free）系统可能是更强有力的实验技术（计算方法除外），可以使我们筛选 10^{14} 个蛋白质变体。显而易见，上述任何一种方法能够处理的蛋白质变体量与理论上可以产生的蛋白质变体量相比是非常局限的，需强调说明必须将库的设计与筛选能力结合起来。实验者或许需计算一个样本中实际蛋白质变体的表征；8.3.1节提供了这样的数学工具。

8.1.3　确定实验偏差

库筛选会产生大量的序列信息（在核酸或者氨基酸水平），必须加以精确阐明。例如，对于库的质量评价需要确定偏差。在库的设计及构建完成后，进行筛选实验前要选出几个样品进行DNA测序。在一个突变位点观察到的与期待的核苷酸分布之间的差异可以用来评价是否存在显著偏差。显著偏差会显示出错误的密码子分布，其来源于寡核苷酸合成失误或者由于野生型序列导致的随机突变设计偏差。偏差也意味着出现了一定量的非期待选择；这样或许需要改变实验系统，如变换表达宿主细胞或者利用体外系统进行表达。阐明实验误差的工具基于 χ^2 检验，将在8.3.2节中详述。

进一步的 χ^2 检验应用还可以比较筛选实验前后的DNA序列来揭示筛选实验造成的密码子分布差别。我们需要观察一个特定残基筛选后的突变频率偏差，或者在一个需要偏好的位点出现特定氨基酸的偏差。应用 χ^2 检验实验者可以确定观察到的偏差是否显著，因此，来确定施加的选择压力对富集某一特定的分布是否足够强。

8.2　材　　料

对于许多计算问题，Excel工作单（图8.1～图8.4）可以提供很好的重复性。通过网页 http：//www.esi.umontreal.ca/～pelletjo/可以直接得到与图中相同的工作单。这些Excel工作单严格遵循此章结构。图8.1和图8.2给出两例等概率和非等概率的计算结果，由于它们是基于相同 λ 参数的泊松分布，因此得到类似的处理。图8.3（再次给出等概率和非等概率的两例）是基于不同 λ 参数的泊松分布。框内单元数据需要用户输入。灰色单元给出答案。

问题A（等概率的）

我们期望有多少理论变量*n*不会出现在所选变量*m*中？

变量数（*n*）	1.00×10^6
样本容量（*m*）	1.00×10^7

期望的丢失变量数（lambda）＝*n**（1-1/*n*）^*m*:　45.40

问题B（等概率的）

在样本中，至少有一个变量不会出现的概率有多少？最多有50个变量不会出现的概率有多少？

至少有一个变量缺失的概率＝1-EXP（-lambda）　100.00%

最多有50个变量不会出现的概率＝POISSON（50；lambda；TRUE）：　77.87%

问题C（等概率的）

在样本中，我们期望一个变量*i*可能出现的次数有多少？更概括讲，变量*i*出现10次的概率有多少？最多出现10次的概率有多少？

变量数（*n*）	1.00×10^6
样本容量（*m*）	1.00×10^7

样本中变量“*i*”发生的期望值（lambda）=m/n:　10.00

样本中变量“*i*”出现10次的概率＝POISSON（10；lambda；FALSE）：　12.51%

样本中变量“*i*”出现少于等于10次的概率＝POISSON（10；lambda；TRUE）：　58.30%

图 8.1　Excel 工作单描述库表征的例子：等概率出现的情形，见 8.3.1.1 1)～3)。

问题A（非等概率的）

我们期望有多少理论变量*n*不会出现在所选变量*m*中？

肽链长度（Pi）	10	← 例如，对十肽来说则用“10”
变量数（n）	1.67×10^{13}	← 仅供参考，不用于此表中
样本容量（m）	1.00×10^{14}	

期望的丢失变量数（lambda）=lambdaPeptide（Pi；m）：5.040×10^{12} ← 框中的VBA函数（lambdaPeptide）是基于密码子数和样本容量这两个参数来计算lambda的函数

问题B（非等概率的）

在样本中，至少有一个变量不会出现的概率有多少？最多有50个变量不会出现的概率有多少？

至少有一个变量缺失的概率＝1-EXP（-lambda）：100.00%

最多有50个变量不会出现的概率=

泊松分布（50；lambda；TRUE）：ERROR(see note) ← 利用泊松分布方程

或返回正态累积分布（50；lambda；SQRT（lambda）；TRUE）：0.00% ← 利用正态分布方程

注：对于许多实际实验数据中得到的非常大值的lambda（如十肽及样本容量*m*为1e14），泊松分布在Excel中的宏很快变为无用值，反馈一个错误信息；在这种情况下，则通过正态分布来处理泊松分布的问题

图 8.2　Excel 工作单描述库表征的例子：非等概率出现的情形，见 8.3.1.2 的 1)和 2)。

问题C（非等概率的）

在样本中，我们期望一个含有7个概率为3/64和3个概率为4/64的密码子的十肽出现的次数有多少？更概括讲，这样的十肽出现10次的概率有多少？最多出现10次的概率有多少？

	"1/64"	"2/64"	"3/64"	"4/64"	"6/64"
概率：	0.015625	0.03125	0.046875	0.0625	0.09375
密码子数	0	0	7	3	0

十肽的概率（p）=
{=PRODUCT（D6:H6^D7:H7）}:　1.21×10^{-13}

样本容量（m）　1.00×10^{14}

样本中变量"*i*"发生的期望值（lambda）=m*p:　12.14

样本中变量"*i*"出现10次的概率=POISSON（10；lambda；FALSE）:　10.23%

样本中变量"*i*"出现少于等于10次的概率=POISSON（10；lambda；TRUE）:　33.27%

注：如果要修改公式p｛=PRODUCT（D6：H6^D7：H7）｝，然后以"ctrl-shift-enter"结束，而不是仅键入"enter"!这是一个整列操作。

图 8.3　Excel 工作单描述一个库表征的实例：非等概率出现的情形，见 8.3.1.2 的 3)。

概率分布的检测（实验偏差的检测）

这里的假设是实验结果反映了理论概率，也就是说实验本身是没有偏差的。

A）2种可能结果

可能结果的数目	2	实验重复的次数	100	显著性水平（“alpha”）	5%
结果1的理论概率	25%	观察到结果1的数目	29	统计值（q）（见*）	0.85
结果2的理论概率	75%	观察到结果2的数目	71	χ^2检验值=CHIINV（L6；1）	3.84
				实验结果是无偏的，则拒绝假设（见**）	Do not reject

B）3种可能结果

可能结果的数目	3	实验重复的次数	20	显著性水平（“alpha”）	5%
结果1的理论概率	20%	观察到结果1的数目	8	统计值（q）（见*）	5.60
结果2的理论概率	30%	观察到结果2的数目	6		
结果3的理论概率	50%	观察到结果3的数目	6	χ^2检验值=CHIINV（L6；1）	5.99
等				实验结果是无偏的，则拒绝假设（见**）	Do not reject

图 8.4　Excel 工作单描述两个概率分布检验的例子：见 8.3.2.1 子节。

* Value of the statistic =(H8-D8*H6)^2/(D8*H6)+(H9-D9*H6)^2/(D9*H6)

** Reject hypothesis? =IF(L8>L10;"Reject";"Do not reject")

8.3 方　　法

此节主要分为两个子节。在 8.3.1 节中，我们提出与库的设计与表征相关的问题。在 8.3.2 节中，我们分析评价库的质量或者库的筛选结果的显著性。在前两节中，我们给出了公式以及提供了数值范例，用示例来加以说明而没有给出详细的理论解释。完整的理论说明可以在注释（见 8.4 节）中找到。

8.3.1　库表征

我们设计一个包含 n 个可能的不同的理论变量库。我们从这个理论库中随机取样 m 次；形成一个 m 个变量的样本集。我们提出如下问题。

A. 在选择的 m 个变量中我们期待多少个理论变量（全部为 n 个）不会出现？

B. 在选择的 m 个变量中至少一个理论变量（全部为 n 个）没有被取样的概率是多少？最多一定量的理论变量（全部为 n 个）没有被取样的概率是多少？

C. 我们期待一个特定变量 i 多少次出现在样本中？更通俗地说，i 出现 r 次的概率是多少？

这些问题作为一组提出，是因为如 8.3.1.1 节和 8.3.1.2 节中所述，上述问题的解决是基于同样的理论。我们用 P_i 表示当我们随机地从理论库中取样一次变量 i 出现任意次的概率。

我们分成两种不同情形，在下面的小节中不同处理：

（1）在 8.3.1.1 节中，等概率结果，即 n 个变量被取样的概率是相同的。这里 $P_i=1/n$，这个例子被作为一般化处理。

（2）在 8.3.1.2 节中，这里结果的出现是非等概率型的，即一些变量比其他一些有更好的概率被取样。根据定义，很难把这种情形进行一般化处理，因为对于上述问题 A、B 和 C，答案取决于概率本身。作为示例，我们详细处理一种非等概率的情形。

如果一个给定序列出现的频率或者我们感兴趣的被编码的特性都没有任何偏差，则称为等概率情形。在这种情形的库选择中，如果我们感兴趣的性质是独立于其他参数的，或者对其他参数的影响可以忽略，我们考虑其为等概率的。如果考虑密码子具有偏好性或简并性，或者结果会影响到整个系统，则为非等概率情形。一个结果影响整个系统的例子如下：当在体内筛选酶 X 的高催化活性时，增加的产物会对整个系统有毒性，产生对检测酶催化活性的不良反应。这样的参数一般来说过去复杂而不用系统地说明。然而我们可以考虑利用偏好性密码子，其中氨基酸的出现是非等概率事件，其他参数尽管也是非等概率的，但是可以被量化。

8.3.1.1　等概率情形

这是最简单的情形，其中所有结果都可以被考虑成等概率事件：对库中的成员进行取样后，所有 n 个变量都有同样的机会被选中，即 $P_i=1/n$。

1）问题 A

在选择的 m 个变量中我们期待多少个理论变量（全部为 n 个）不会出现？

一个简单近似的答案由下面参数 λ（详见注释 1 及 2）定义给出：

$$\lambda=n\left(1-\frac{1}{n}\right)^{m} \tag{8.1}$$

例如，假设我们从 $n=1\times10^6$ 个理论变量库中取样 $m=1\times10^7$ 次。任意一个变量出现任意一次的概率是 $p_i=\frac{1}{n}=\frac{1}{1\times10^6}$。

缺失掉的理论变量的数目的期望值由参数 λ 计算出，如下：

$$\lambda=1\times10^6\left(1-\frac{1}{1.0\times10^6}\right)^{1\times10^7}\approx45.4$$

注意，缺失变量数目的期望值不必是整数；这里缺失变量数目的期望值为 45～46 个（整体理论变量数目为 1×10^6 个）。

如果我们从 $n=1\times10^6$ 个理论变量库中取样 $m=2\times10^7$ 次，则 $\lambda=0.002\ 06$。这样，缺失变量的期望值为 0 或 1 个，且更加接近于 0。事实上，在这个例子中，0 或 1 个变量缺失的概率是仅有的非小概率事件，0 缺失更加可能（>99%，见 8.3.1.1 问题 B 中的计算）。

我们可以得出结论：取样大小为理论库的 20 倍时（$m=20n$），可以导致原则上对理论库的完全取样，如图 8.5 所示。取样 10 倍过量可以导致一小部分组分没有被取样（当理论库中少于 10^4 个变量时，此组分趋向于 0，见表 8.2）；对于很多的实验目的来说，知道还有一小部分没有被取样已经达到要求。

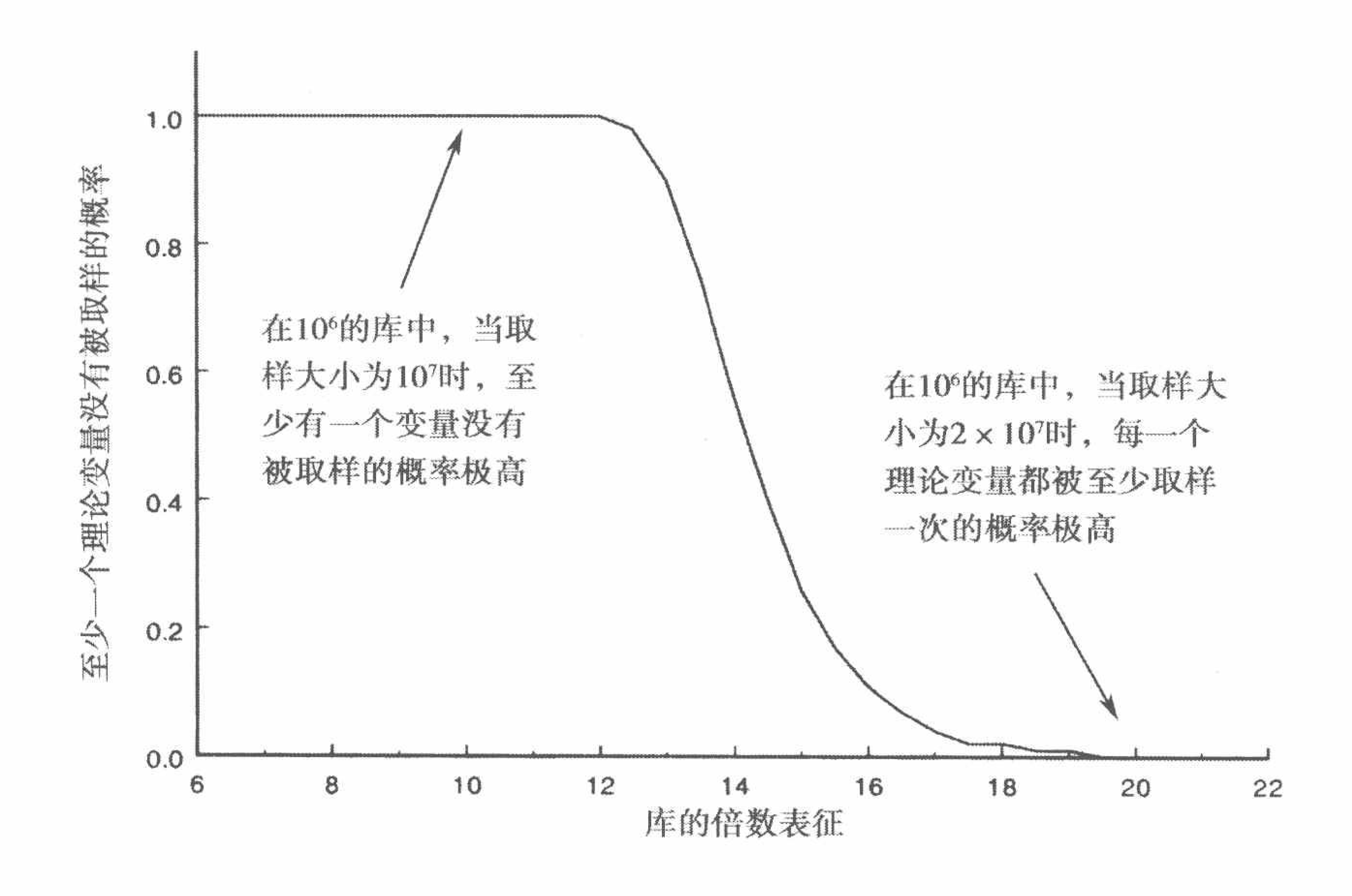


图 8.5 至少一个理论变量没有被取样的概率相对于库的倍数表征。理论库的大小为 n，取样数为 m；库的表征为比率（倍数）n/m，10 倍库表征意味着 $m=10n$。这里的示例限定了理论库大小 $n=1\times10^6$ 的情形，m 也随之被限定。

表 8.2

n	m	库的倍数表征（n/m）	未被取样的期望值（λ）	未被取样的期望百分率/%	P（至少有1个变量缺失的概率）/%
1000	500	0.5	606	60.6	100
	1000	1	368	36.8	100
	2000	2	135	13.5	100
	10 000	10	0.05	0.00	4.42
	20 000	20	0.00	0.00	0.00
1×10^6	5×10^5	0.5	6.07×10^5	60.7	100
	1×10^6	1	3.68×10^5	36.8	100
	2×10^6	2	1.35×10^5	13.5	100
	1×10^7	10	45.4	0.00	100
	2×10^7	20	0.00	0.00	0.21
1×10^{10}	5×10^9	0.5	6.07×10^9	60.7	100
	1×10^{10}	1	3.68×10^9	36.8	100
	2×10^{10}	2	1.35×10^9	13.5	100
	1×10^{11}	10	4.45×10^5	0.00	100
	2×10^{11}	20	20.6	0.00	100

请注意表 8.2 给出一些违反直觉的结果：如果取样大小 m 等于库的大小 n，我们可以期待平均大概 37%的理论变量将不被取样；甚至当 $m=2n$ 时，理论变量不被取样的平均概率是 13.5%，远小于应该被完全取样的结果。

2）问题 B

在选择的 m 个变量中至少一个理论变量（全部为 n 个）没有被取样的概率是多少？最多一定量（50 或 60）的理论变量（全部为 n 个）没有被取样的概率是多少？

k 个变量缺失的近似概率由下式给出：

$$P（k\text{个变量缺失}）\approx\frac{e^{-\lambda}\lambda^{k}}{k!},\ k\in\{0,1,\cdots,n\} \tag{8.2}$$

式中，λ被公式（8.1）定义（参见注 1 和注 3）。从这里可以看出，至少一个理论变量缺失的概率大约是 $1-e^{-\lambda}$（参见注 4）。最多 K 个理论变量缺失的概率可由 $k=0$，1，2，…直到 K 的概率值加和［见公式（8.2）］得到。

要解决问题 B，我们需要首先如 8.3.1.1 小节 1）中算出 λ 值。在问题 A 的示例中，我们从1×10^6 个变量的理论库中取样 1×10^7 次，可以算出 $\lambda\approx45.4$。然后，至少一个变量没有被取样的概率计算为 $1-e^{-45.4}\approx1.00$ 或 100%。因此，肯定（100%）至少有一些理论变量在 m 个取样中丢失掉了。在实际中，这意味着什么呢？这一结果（以及更多的一些例子见表 8.2）指示如果 n 等于或大于 10^5，筛选取样数 m 大于 10 倍的库变量数 n，则非常可能（几乎是 100%）理论库 n 没有被完全取样。对于小的库（$n=1000$ 或更小），则 $m=10n$ 给出不完全覆盖理论库的概率是很小的（如对于 $n=1000$，概率仅约为 5%，见表 8.2）。

相应地，如果我们取样多一倍，即 $m=2\times10^7$ 取样同样的理论库 $n=1\times10^6$ 个变量，则 $\lambda=0.0026$，并且至少一个理论变量没有被取样的概率大约是 $1-e^{-0.0026}=0.002\ 06$ 或者 0.2%。在这个例子里，所有理论变量 n 都被取样的概率大于 99%（更加精确地说是

100%−0.206%=99.794%）。事实上，λ和概率 $1-e^{-\lambda}$ 相近是不奇怪的，当 x 很小时，$e^x\approx 1+x$）。

对于我们的主要例子（$n=1\times 10^6$；$m=2\times 10^7$），从 n 个理论变量中选取的 m 次中少于 50 个没有被取样的概率可以由加和每一个 0，1，2，…，直到 49 个没有被取样的概率计算得出，大约是 78%，即

$$\sum_{k=0}^{49} e^{-0.00206}\frac{0.00206^k}{k!}=0.78$$

与此类似，少于 60 个没有被取样的概率大约是 98%。为了简化计算，可以利用 Excel 计算单（见 8.2 节及图 8.1）。

3）问题 C

我们能够期待一个特定变量 i 在取样中出现多少次？1 个变量 i 出现 10 次的概率是多少？3 次呢？更普遍的说，变量 i 出现 10 次和 10 次以下的概率是多少？

一个变量在取样中出现特定次数的概率可由泊松分布（见注 5）近似给出，利用参数

$$\lambda=\frac{m}{n} \tag{8.3}$$

则我们期待一个特定变量 i 出现的次数是

$$\text{期待变量 } i \text{ 出现的次数}=\lambda=\frac{m}{n}$$

并且概率为

$$\text{概率 } P\text{（变量 } i \text{ 在 } m \text{ 大小的取样中出现 } r \text{ 次）}=e^{-\lambda}\frac{\lambda^r}{r!} \tag{8.4}$$

（见注 2、注 3 和注 6）。注意在这个问题中由方程（8.3）给出的泊松分布的参数是不同于问题 A 和问题 B 用到的参数，见子节 3.1.1.1 及 3.1.1.2 及方程（8.1）。

一个变量出现特定次数及以下的概率可由方程（8.4）重复地计算出。

例如，我们再用 $n=1\times 10^6$ 和 $m=1\times 10^7$。我们期待在取样中可以找到变量 i $\frac{1\times 10^7}{1\times 10^6}=10$ 次。i 出现 10 次的概率是

$$e^{-\lambda}\cdot\frac{\lambda^r}{r!}=e^{-\frac{1\times 10^7}{1\times 10^6}}\cdot\frac{\left(\frac{1\times 10^7}{1\times 10^6}\right)^{10}}{10!}\approx 12.5\% \tag{8.5}$$

利用嵌入在一些软件程序包（如 Excel、Matlab、Mathematica 等）中的泊松分布函数，上述计算可以被轻松完成。例如，可以参见附于此章的 Excel 计算单（见 8.2 节及图 8.1）。

如图 8.6 所示，得到 10 次同一变量的可能性是蛮好的（12.5%），计算旁边的其他情况（i=8、9、11 和 12）可以分别给出得到同一变量相应次数的概率也是挺高的（分别是 11.3%、12.5%、11.4%和 9.5%）。然而，变量 i 出现 3 次的概率仅为

$$e^{-\frac{1\times 10^7}{1\times 10^6}}\cdot\frac{\left(\frac{1\times 10^7}{1\times 10^6}\right)^{3}}{3!}\approx 0.76\%$$

即小于 1%的机会。某个变量出现 3 次的结果会提示系统中有偏差吗（如由于选择压力造成的）？关于这个问题的讨论见 8.3.2 节。

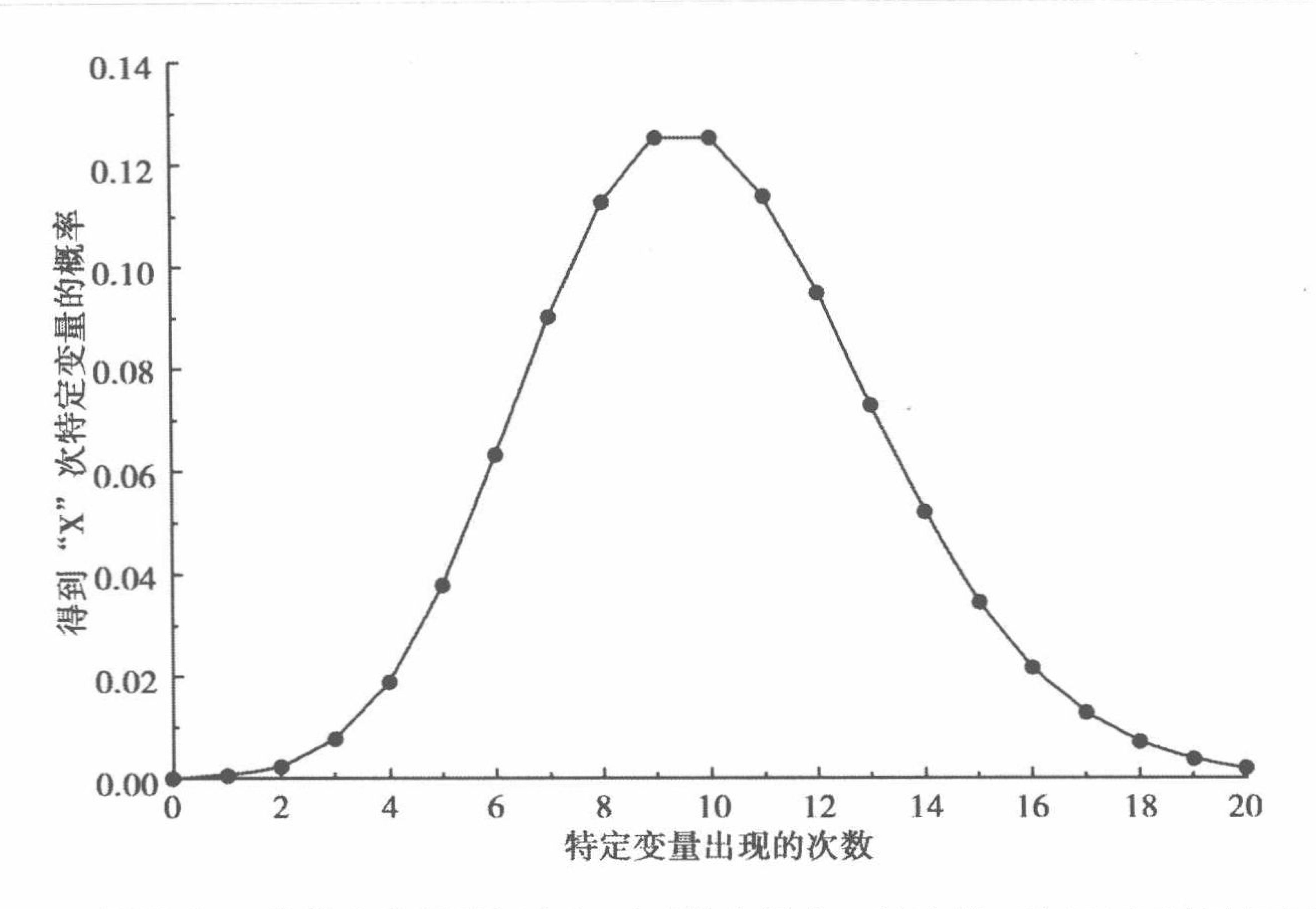


图 8.6　得到“*x*”次特定变量的概率相对于此变量出现的次数。这里展示的例子是使用的理论库大小 $n=1\times10^6$ 和取样数 $m=1\times10^7$。这样得到 8、9、10 或 11 次特定变量的概率都大于10%，得到 8～12 次的概率大于 50%。

最后，某变量出现 10 次或以下的概率可以由相加出现 10 次，9 次等的概率直到出现 0 次的概率算出，可以手算也可以利用 Excel 工作单（见 8.2 及图 8.1）。我们看到在这个例子中出现 10 次或更少的概率大约是 58.3%。

8.3.1.2　非等同概率的一个例子

我们给出出现非等同概率事件的一个例子，即 *n* 个变量中的一些变量会比平均情况更多或更少的机会出现。如 8.3.1 节开头所述，很难得出非等同概率事件的普遍规律，因为答案取决于具体概率是什么。我们这里提供的例子中给出一些简明的计算，同时也提示了可能遇到的困难。

在没有选择压力时，非等同概率事件可以在有偏好的 DNA 库中出现，这种库偏好性可以是设计的也可以是意外产生的。例如，在寡核苷酸引物合成过程中出现的意外偏差。这样的偏差可以在随机取样 DNA 库测序后发现。如果我们用编码蛋白质序列而不是 DNA 序列来定义理论库，那么非等同概率依然会出现。例如，当“NNN”型密码子编码时，将会出现 6 个编码亮氨酸的每一个密码子出现的概率与唯一编码甲硫氨酸密码子出现的概率相等，因此在同一位点得到含有亮氨酸的蛋白质的概率将是得到含有甲硫氨酸的蛋白质概率的 6 倍。

在选择压力存在时，实验开始时一般来说不可能精确数值定义出某些突变相关的选择优势或者劣势。事实上，在有选择压力时最好反过来问问题：与其问给出特定的序列偏好，某种变种以何种概率出现，不如问在选择出的子集中我们能够观察到的序列的实验频率。我们能够确定这样的频率是否为期待的如 8.3.1.1 3 中的问题 C)，或者是否出现了偏差。

1）我们这个非等同概率例子的特征

我们考虑在十肽中取样。如果 10 个氨基酸在十肽中是随机分布的（NNN），会有

$21^{10}\approx1.67\times10^{13}$种不同的理论变种（我们对待终止子与其他氨基酸相同）。因为密码子的简并性每种氨基酸以非等同概率被选择；尽管如此，能够发生的不同概率为下列 5 种：

$$\left\{\frac{1}{64},\frac{2}{64},\frac{3}{64},\frac{4}{64},\frac{6}{64}\right\}$$

其中 Met 和 Trp 分别被 64 个密码子中的一个编码；Phe、Tyr、His、Gln、Asn、Lys、Asp、Glu 和 Cys 分别被 64 个密码子中的两个编码等。因此，上述 5 种概率所对应的氨基酸种类，分别为 2、9、2、5 以及 3 种氨基酸或终止密码子。

2）问题 A

在 n 个理论变种中选出 m 个，我们可以预计多少个不出现？

要解决这个问题，要素是考虑密码子的简并性。我们假定每个特异的十肽无论其氨基酸排列顺序如何，都具有与其他相同氨基酸组成（因此具有同样的简并性）的十肽相同的出现概率。则问题 A 的答案是参数 λ（见注 7），计算式为

$$\lambda\approx\sum_{(n_1,n_2,n_3,n_4,n_5)}\left[\frac{10!}{n_1!\,n_2!\,n_3!\,n_4!\,n_5!}2^{n_1}9^{n_2}2^{n_3}5^{n_4}3^{n_5}\right]e^{-p(n_1,n_2,n_3,n_4,n_5)m} \tag{8.6}$$

上述加和号是对所有 5 个数（n_1、n_2、n_3、n_4、n_5）组成的数组的不同排序求和，相应于给定密码子简并性的 5 种概率。每个取自集合 {0，1，2，…，10} 的“5 数组”代表十肽中 5 种不同简并度的氨基酸各占多少。因为我们只考虑 10 种氨基酸，所以，$\sum_{i=1}^{5}n_i=10$。例如，(n_1，n_2，n_3，n_4，n_5) ＝（8，0，0，1，1）意味着一个十肽中的 8 个位置是由非简并氨基酸（1/64）编码，不包含简并度为 2/64 和 3/64 的氨基酸，10 个位置中简并度为 4/64 和 6/64 的氨基酸各有一个。

p（n_1，n_2，n_3，n_4，n_5）可以计算为

$$p(n_1,n_2,n_3,n_4,n_5)=\left(\frac{1}{64}\right)^{n_1}\left(\frac{2}{64}\right)^{n_2}\left(\frac{3}{64}\right)^{n_3}\left(\frac{4}{64}\right)^{n_4}\left(\frac{6}{64}\right)^{n_5} \tag{8.7}$$

这个 p（n_1，n_2，n_3，n_4，n_5）是我们在 8.3.1 节开头简要介绍的 p_i 的一个新名字，意味着当我们从理论库中随机取样一次，特定蛋白变种出现任意次的概率。

不幸的是，对于参数 λ 似乎不存在更简单的表述；要计算 λ 的求和必须利用计算机编程。

例如，我们要从理论库 $n=21^{10}$（即约 1.67×10^{13}；见图 8.2）中随机取出 $m=10^{14}$ 个十肽。

对于这组 m 和 n 的答案是 $\lambda=5.04\times10^{12}$，因此我们预计从 1.67×10^{13} 的理论库中随机取样有 5.04×10^{12} 个不出现，大概的比率是 30%。

3）问题 B

样本中至少有一种多肽变种不被产生出来的概率是多少？最多 50 或 60 种多肽变种没有被取样的概率是多少？

这样的概率可以由 8.3.1.1 节中有关等概率事件方程（8.2）得到；但是参数 λ 的值必须由方程（8.6）计算。因此，至少一种多肽不被取样的概率可以近似为（见注 5）：

$$1-e^{-\lambda} \tag{8.8}$$

计算“至少”或者“最多”特定数目多肽不被取样的概率可由 8.3.1.1 2）的方法得出，这里不再赘述。

例如，我们再次从理论库 $n=21^{10}\approx1.67\times10^{13}$ 中随机取出 $m=10^{14}$ 个十肽；并且在

8.3.1.1 的子节 2）中求得 $\lambda=5.04\times10^{12}$。显然，对于 $\lambda=5.04\times10^{12}$ 的预计缺失变种，直觉认为至少缺失一个的概率应该是很高的。利用式（8.8），我们确实可以求出此概率是 100%。换句话说，可产生出所有蛋白质变种的概率是 0%。

事实上，在所有多肽变种中最多 50 或 60 种不被产生的概率也是 0%。对于这样的高 λ 值（5.04×10^{12}），这样的结果并不奇怪。

4）问题 C

在一个样本中，我们可以预计含有 7 个概率为 3/64 的密码子和 3 个概率为 4/64 的密码子的十肽会出现多少次？这样的十肽出现 10 次的概率是多少？3 次呢？出现 10 次及 10 次以下的概率呢？

因为对于所有的在这个问题中所定义构成的肽来说，结果将会是一样的，所以我们没有必要去关心确切的序列（比如氨基酸的次序和名称）。这些问题的答案只是在 8.3.1.1 的 3）中给出的问题中有微小的差别。因为对任意的十肽进行抽样后一个十肽出现的概率为 p［见 8.3.1.1 的 2)］，这样我们所预计的十肽出现的次数就是 mp，这里 m 是抽样的次数。考虑到一些十肽会被抽样超过一次，这样概率将会随着抽样次数而增加，所预计的十肽出现一定次数的概率近似服从参数为 $\lambda=mp$ 的泊松分布（参见注 3 和注 8），就是

$$\text{概率（十肽在样本容量为 } m \text{ 的样本中出现 } r \text{ 次）}=e^{-\lambda}\frac{\lambda^r}{r!}$$

举个例子，再次假设我们对一个 $m=10^{14}$ 样本容量的样本抽样，这个十肽出现的概率（很小）为

$$p(0,0,7,3,0)=\left(\frac{3}{64}\right)^7\left(\frac{4}{64}\right)^3=1.21\times10^{-13}$$

也就是说，我们可以期望在样本中找到 $\lambda=mp=10^{14}\times(1.21\times10^{-13})\approx12.14$ 次这个十肽。出现十次的概率为

$$e^{-\lambda}\frac{\lambda^r}{r!}=e^{-12.14}\left(\frac{12.14^{10}}{10!}\right)\approx10.2\%$$

同样的，特定的十肽出现 3 次的概率近似为 0.16%，远远小于出现 10 次的机会（10%）。把小于 10 次的概率全部加起来，我们得到出现小于或等于 10 次的概率近似为 33%（图 8.3）。

8.3.2　实验偏差的检验

这一节的基本问题可以表述为具有两个或两个以上可能观测值的试验进行了多次，也就是，多次取样的问题。每个观测值的理论概率（预期的）是已知的，而观测值的试验概率可以通过观察得到。理论概率和试验概率间的差别是否是显示实验有显著性偏差（biased）的信号？

如节 8.1 中被提及的，在选择之前或者之后，这样的分析可以被常规地应用到 DNA 测序结果中测试偏差。应用下面提到的 χ^2 检验可以让试验者确定观察到的偏差是显著的还是不显著的。

现在由以下记号来提出问题：一个试验在理论上得到 k 个观测值 A_1，A_2，…，A_k 的概率分别为 p_1，p_2，…，p_k。例如，对一个蛋白质库取样可以得到 $k=2$ 个结果：有功能或非功能的。这个试验被做了 n 次（也就是，n 个蛋白质被随机地抽样），则试验结果在每个观测值中出现的次数分别是 Y_1，Y_2，…，Y_k 次。

χ^2检验

χ^2检验中的统计量为

$$q=\frac{(Y_1-np_1)^2}{np_1}+\frac{(Y_2-np_2)^2}{np_2}+\cdots+\frac{(Y_k-np_k)^2}{np_k}$$

我们要检验的假设，被称作“零假设”，是试验以合适方式进行时的平均值，也就是说，以这一方式所得各个观测值的概率符合它们的理论概率。一个大的 q 值表明真实的试验结果与期望的结果之间存在显著的差异，而与零假设矛盾。

把统计量 q 与参数为 $k-1$ 在任意选择的显著性水平 α 下得到的值χ^2_α比较。参数 α 的值由使用者指定（关于显著性水平见注 9）。检验如下：

当 $q<\chi^2_\alpha(k-1)$ 时，接受原假设；没有充分的理由认为存在偏差。

当 $q>\chi^2_\alpha(k-1)$ 时，拒绝原假设；有充分的理由认为存在偏差。

χ^2分布的临界值可以在大部分的统计学课本和很多软件，如 Excel 中查到。

例 1 两个可能的观测值：一个实验被进行了 100 次；每一次一个特定的特性出现或是不出现，因此，现在有两个可能的观测值并且 $k=2$。理论上，这个特性（称为“观测值 1”）应该在 4 次抽样中出现 1 次；在 100 次抽样中，它出现了 29 次。χ^2检验基于统计量

$$q=\frac{(29-100\times0.25)^2}{100\times0.25}+\frac{(71-100\times0.75)^2}{100\times0.75}\approx0.85$$

让我们选择一个显著性水平 $\alpha=0.05$。则比较 $q=0.85$ 与

$$\chi^2_{0.05}(k-1)=\chi^2_{0.05}(2-1)=\chi^2_{0.05}(1)=3.84$$

因为：

$q=0.85<3.84=\chi^2_{0.05}(1)$

原假设（试验以无偏好的正确方式进行）被接受（图 8.4）。图 8.7 描述了该假设具有应该被接受（17～33 个观测值的特征）还是被拒绝（所有其他结果）的特征的观测值个数。

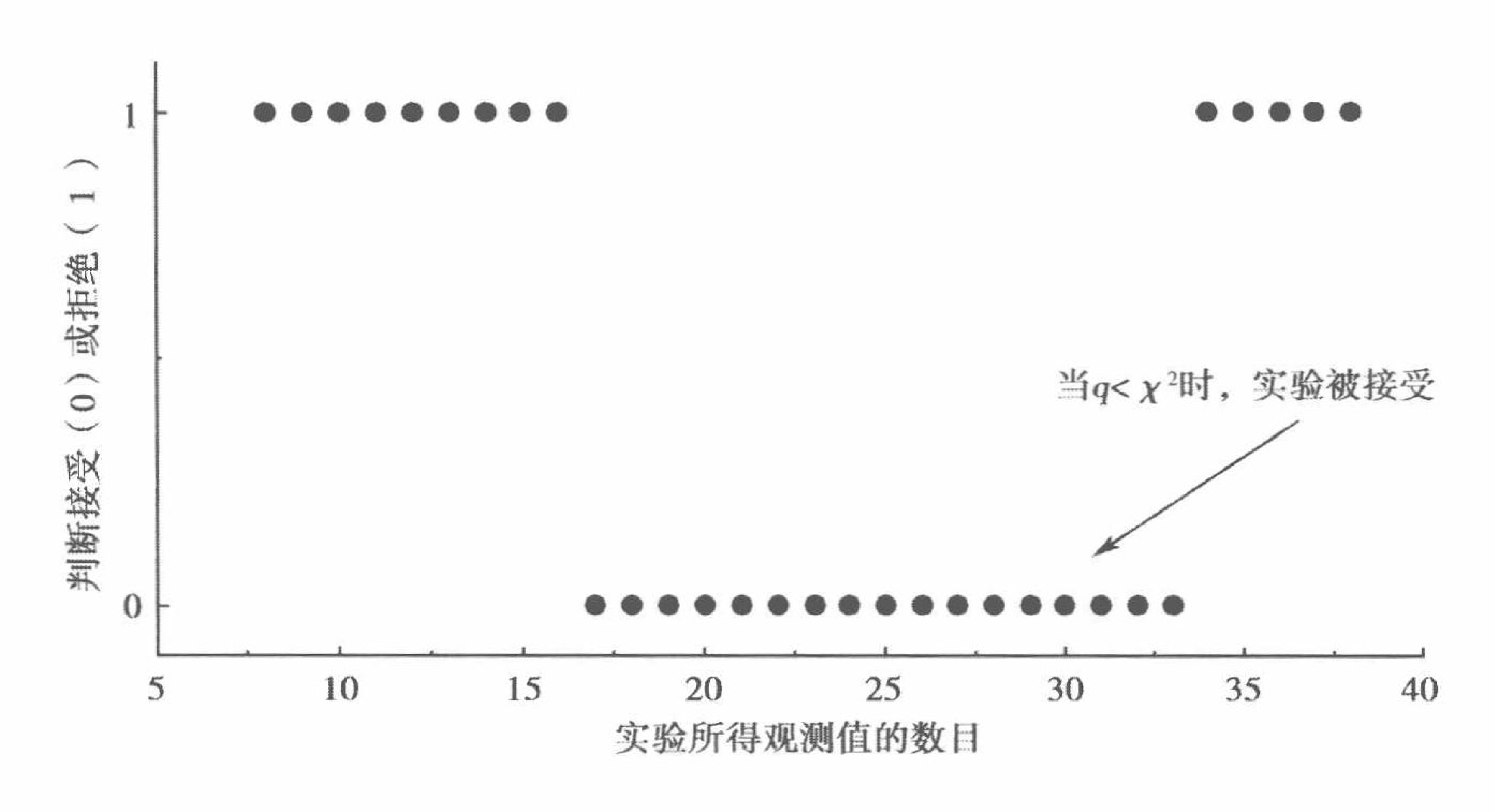


图 8.7 根据χ^2检验判断接受还是拒绝一个试验结果。对于一个试验观测值计算它的 q 值，并与根据要求选择的显著性水平 α 查表得到的χ^2检验值进行比较。如果 $q<\chi^2$，观测值的值被置为 0，被接受。如果 $q>\chi^2$，观测值的值被置为 1，被拒绝。

例 2　三种可能的观测值：设理论上一个试验产生观测值 A 的可能为 20%；观测值 B 的可能为 30%；以及观测值 C 的可能为 50%。让一个实验室的新同学重复这一试验 20 次；这一学生得到观测值 A 有 8 次，观测值 B 有 6 次，以及观测值 C 6 次。这是可信的吗？这个学生的技术是可信赖的吗？

这一问题的统计量是

$$q=\frac{(8-20\times 0.20)^2}{20\times 0.20}+\frac{(6-20\times 0.30)^2}{20\times 0.30}+\frac{(6-20\times 0.50)^2}{20\times 0.50}\approx 5.6$$

在显著性水平 α 为 5%时，我们得到 $\chi^2_{0.05}(2)=5.991$，由

$$q=5.6<5.991=\chi^2_{0.05}(2)$$

表明统计量 q，以及对期望结果的不符合性不足够大，因此我们可以得出结论认为在这个学生的操作步骤中存在瑕疵；而原假设没法被拒绝。然而，如果我们取 $\alpha=10\%$，则 $\chi^2_{0.10}(2)=4.605$，检验变成了：$q=5.6>4.605=\chi^2_{0.10}(2)$ 这样，我们能够拒绝原假设，认为现在的试验是所期望的。

8.4　注

1. 对于问题 A 和 B 的解答 8.3.1.1 的 1）和 2）依赖于伯努里（Bernoulli）随机变数求和的泊松近似理论，这一理论主要是说，如果所有在方程（8.9）和方程（8.14）中定义的 λs 值都很小，那么没有在样本中出现的变量的个数将近似服从参数为 $\lambda=\sum_{i=1}^{n}\lambda_i$ 的泊松分布［关于数学具体算法请参见参考文献［15］的第 10 章，特别是关于多项分布的例 10.2（B）］。

根据参考文献［16］的第 252 页，泊松近似在 $n\geqslant 20$，$\lambda\leqslant 10$ 时效果最佳。

2. 没有出现的变量个数近似服从参数为 $\lambda=\sum_{i=1}^{n}\lambda_i$ 的泊松分布，这里，λ_i 可被定义为

$$\lambda_i=P（变量\ i\ 是在样本中没有出现的）=(1-p_i)^m \tag{8.9}$$

根据泊松分布的性质，没有出现的变量的个数的期望值就是参数 λ 本身。

在等概率事件的情况下，因为 $p_i=\frac{1}{n}$，这样可以简单算出 λ_i 和 λ

$$\lambda_i=(1-p_i)^m=\left(1-\frac{1}{n}\right)^m$$

以及：

$$\lambda=\sum_{i=1}^{n}\left(1-\frac{1}{n}\right)^m=n\left(1-\frac{1}{n}\right)^m \tag{8.10}$$

注意，给出特定的 λ_i，λ 值可被直接计算而无须计算 λ_i。

3. 用 Excel 计算泊松分布在分布系数 λ 很大时依然适用。在这种情况下，可以简单地运用均值等于 λ 和方差等于 λ 的正态分布，也就是 $N(\lambda, \lambda)$ 替代。这一技巧的举例说明见图 8.2。

4. 由于泊松分布的特点

$$P（k\ 个变量没有出现）\approx\frac{e^{-\lambda}\lambda^k}{k!}，k\in\{0, 1, \cdots, n\} \tag{8.11}$$

这里 λ 在式（8.1）中或者在式（8.10）中被定义。

没有出现的个数为零时的概率近似为

$$\frac{e^{-\lambda}\lambda^0}{0!}=e^{-\lambda}$$

因此至少一个没有出现的概率可近似为

$$1-e^{-\lambda} \tag{8.12}$$

5. 问题 C 的解答是基于次数计数这一随机变量服从变量的二项分布的泊松近似。这里，根据参考文献［16］，泊松近似在 $n\geqslant 20$，$\lambda\leqslant 10$ 时效果最佳。

6. 在样本中得到一个特定变量 i 的事件为一定次数 r 的概率服从参数为 m（样本容量）和 $1/n$（选取变量 i 的概率，任何时刻一个变量被选取是随机的）的二项分布。这一二项分布可以很好地近似为参数为 $\lambda=\frac{m}{n}$ 的泊松分布，而且计算起来也更容易。

7. 见注 2。对于这种情况，λ 可以计算如下。不同于等概率情况，λ_i 在这里是不同的。首先，为了计算 λ，我们需要计算所有的 $n=21^{10}$ 个参数 λ_i，这个工作量很大，但是这里包含了很多重复，这一特征可以用来简化计算。而且，让我们确定一些由五元组（n_1，n_2，n_3，n_4，n_5）确定的十肽，由密码子的数量可知，它们的概率在所描述的十肽中分别为 $\frac{1}{64}$，$\frac{2}{64}$，$\frac{3}{64}$，$\frac{4}{64}$ 和 $\frac{6}{64}$。必然 $\sum_{i=1}^{5}n_i=10$，否则在十肽中不会有 10 个密码子了。并且有

$$\frac{10!}{n_1!\,n_2!\,n_3!\,n_4!\,n_5!}2^{n_1}9^{n_2}1^{n_3}5^{n_4}3^{n_5} \tag{8.13}$$

种不同的十肽与五元组（n_1，n_2，n_3，n_4，n_5）有关。这一数字解释了多肽中密码子的所有混编，也就是多肽中的氨基酸次序不需要被明确定义。它同样说明了这样一个事实，不止一个氨基酸能表现出其中的每一个概率（除了异亮氨酸，异亮氨酸是唯一一个被 3/64 密码子编码的氨基酸）。我们提出的观点是，五元组包含了用来回答我们问题的关于十肽的所有信息。现在我们可以用 λ（n_1，n_2，n_3，n_4，n_5）来替换 $\lambda_i=(1-p_i)^m$ 的计算，因此，对于特定的十肽 i：

$$\lambda_i=\lambda(n_1,n_2,n_3,n_4,n_5)=[1-p(n_1,n_2,n_3,n_4,n_5)]^m \tag{8.14}$$

这对于与十肽 i 相关的五元组（n_1，n_2，n_3，n_4，n_5）是特定的。也就是说，对于 21^{10} 个参数 λ_i，我们只需要计算 λ_i 中出现的不同值。λ_i 中一共有

$$\binom{10+5-1}{5-1}=\frac{14!}{10!\,4!}=1001 \tag{8.15}$$

个不同的值；这里，我们用了二项系数记号，$\binom{r}{s}=\frac{r!}{(r-s)!s!}$。举个例子，见参考文献［17］的箱中球分布问题部分，有关于式（8.15）的细节描述。

因为描述服从一个多项分布，相关于五元组（n_1，n_2，n_3，n_4，n_5）的十肽 i 出现的概率 p_i 为

$$p_i=p(n_1,n_2,n_3,n_4,n_5)=\left(\frac{1}{64}\right)^{n_1}\left(\frac{2}{64}\right)^{n_2}\left(\frac{3}{64}\right)^{n_3}\left(\frac{4}{64}\right)^{n_4}\left(\frac{6}{64}\right)^{n_5} \tag{8.16}$$

因此可以很简单地通过式（8.14）计算 $\lambda_i=\lambda$（n_1，n_2，n_3，n_4，n_5）了。

我们最后只用这 1001 个 λ（n_1，n_2，n_3，n_4，n_5）值来计算 $\sum_{i=1}^{21^{10}}\lambda_i$，以及每个数值根

据式（8.13）所给出的重复次数。

$$\lambda = \sum_{i=1}^{21^{10}} \lambda_i$$
$$= \sum_{(n_1,n_2,n_3,n_4,n_5)} \left| \frac{10!}{n_1!n_2!n_3!n_4!n_5!} 2^{n_1} 9^{n_2} 1^{n_3} 5^{n_4} 3^{n_5} \right| \lambda(n_1, n_2, n_3, n_4, n_5)$$
$$= \sum_{(n_1,n_2,n_3,n_4,n_5)} \left| \frac{10!}{n_1!n_2!n_3!n_4!n_5!} 2^{n_1} 9^{n_2} 1^{n_3} 5^{n_4} 3^{n_5} \right|$$
$$\times [1 - p(n_1, n_2, n_3, n_4, n_5)]^m$$
$$= \sum_{(n_1,n_2,n_3,n_4,n_5)} \left| \frac{10!}{n_1!n_2!n_3!n_4!n_5!} 2^{n_1} 9^{n_2} 1^{n_3} 5^{n_4} 3^{n_5} \right| e^{-p(n_1,n_2,n_3,n_4,n_5)m}$$

这里第二，第三和第四步的求和都是定义在 8.1 中的五元组（n_1，n_2，n_3，n_4，n_5）。也就是说，它们都是五套数字，且满足$\sum_{i=1}^{5} n_i = 10$，这里 n_i是在 {0，1，…，10} 中的个数，因为它是一个十肽。同样的，p（n_1，n_2，n_3，n_4，n_5）根据式（8.16）计算得到。最后一个方程给出了 λ 的近似值（对于很小的 x，$1+x \approx e^x$），这个运算要明显比从第二个方程到最后一个方程简单；通常，在计算机上计算，$1-p$（n_1，n_2，n_3，n_4，n_5）将趋近于 1。

8. 见注 5。这里唯一的不同是二项分布的基本事件概率不再是 $1/n$（等概率部分描述的概率），而是其他的一些基于不同的变量的概率 p。

9. 回顾显著性水平 $\alpha = 0.05$ 的意义。这表示有 5%的机会犯第一类错误，意味着假设为真却被拒绝。换句话说，这里有 1∶20 的机会说明一个完美执行的试验有缺陷。

致谢　我们对 Andreas Plückthun 和他的小组所提出的问题表示感谢，除此之外，我们还要感谢 Bruno Rémillard（HEC 蒙特利尔，加拿大），Nicolas Doucet（蒙特利尔大学，加拿大），Sabine C. Stebel（弗莱堡大学，德国）以及本书的编辑 Katja M. Arndt 和 Kristian M. Müller（弗莱堡大学，德国）所提出的很有价值的意见。

本工作得到了国家自然科学和工程研究委员会项目 227853-02（J. N. Pelletier）和 227838-00（M. Denault）的资助。

参考文献

1. Arnold, F. H. and Georgiou, G. (2003) *Directed Evolution Library Creation Methods and Protocols.* Methods in Molecular Biology 231, Humana Press, Totowa, NJ.
2. Arnold, F. H. and Georgiou, G. (2003) *Directed Enzyme Evolution Screening and Selection Methods.* Methods in Molecular Biology 230, Humana Press, Totowa, NJ.
3. Moore, J. C., Jin, H. M., Kuchner, O., and Arnold, F. H. (1997) Strategies for the *in vitro* evolution of protein function—enzyme evolution by random recombination of improved sequences. *J. Mol. Biol.* 272, 336-347.
4. Moore, G. L. and Maranas, C. D. (2000) Modeling DNA mutation and recombination for directed evolution experiments. *J. Theor. Biol.* 205, 483-503.
5. Moore, G. L., Maranas, C. D., Lutz, S., and Benkovic, S. J. (2001) Predicting crossover generation in DNA shuffling. *Proc. Natl. Acad. Sci. USA* 98, 3226-3231.
6. Moore, G. L. and Maranas, C. D. (2002) Predicting out-of-sequence reassembly in DNA shuffling. *J. Theor. Biol.* 219, 9-17.
7. Sun, F. (1999) Modeling DNA shuffling. *J. Comput. Biol.* 6, 77-90.
8. Patrick, W. M., Firth, A. E., and Blackburn, J. M. (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng.* 16, 451-457.

9. Bittker, J. A., Le, B. V., Liu, J. M., and Liu, D. R. (2004) Directed evolution of protein enzymes using nonhomologous random recombination. *Proc. Natl. Acad. Sci. USA* 101, 7011-7016.

10. Gribskov, M., Devereux, J., and Burgess, R. R. (1984) The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. *Nucleic Acids Res.* 12, 539-549.

11. Virnekas, B., Ge, L., Plückthun, A., Schneider, K. C., Wellnhofer, G., and Moroney, S. E. (1994) Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.* 22, 5600-5607.

12. Pelletier, J. N., Arndt, K. M., Plückthun, A., and Michnick, S. W. (1999) An *in vivo* library-versus-library selection of optimized protein-protein interactions. *Nature Biotechnol.* 17, 683-690.

13. Braunagel, M. and Little, M. (1997) Construction of a semisynthetic antibody library using trinucleotide oligos. *Nucleic Acids Res.* 25, 4690, 4691.

14. Gaytan, P., Yanez, J., Sanchez, F., and Soberon, X. (2001) Orthogonal combinatorial mutagenesis: a codon-level combinatorial mutagenesis method useful for low multiplicity and amino acid-scanning protocols. *Nucleic Acids Res.* 29, E9.

15. Ross, S. M. (1996) *Stochastic Processes.* 2nd ed., John Wiley & Sons, New York, NY.

16. Hogg, R. V. and Tanis, E. A. (1983) *Probability and Statistical Inference.* 2nd ed., MacMillan, New York, NY.

17. Ross, S. M. (1998) *A Fist Course in Probability.* 5th ed., Prentice Hall, Upper Saddle River, NJ.

9 基于极性与非极性氨基酸的“二元组图”进行蛋白质设计

Luke H. Bradley, Yinan Wei, Peter Thumfort, Christine Wurth, and Michael H. Hecht

概要 设计具有很好折叠的大规模新型蛋白质库是最终能够产生具有工业及医学应用新结构及功能的蛋白质的有效途径。本章介绍一个极性/非极性氨基酸“二元组图”的蛋白质库设计方法，此方法整合了合理化设计和重组多样性。“二元组图”基于的前提是极性/非极性氨基酸残基的适当组合可以指导多肽链折叠成具有双亲性质的二级结构，此类二级结构可能“退火”形成所需的高级结构。设计的二元组图利用蛋白质二级结构中本身具有的周期性，使得每一个极性和非极性位点的侧链可以产生的组合变化。本章概述了如何利用二元组图法设计新型蛋白质库。

关键词 蛋白质设计；二元组图；组合库；新型蛋白质；蛋白质库设计

9.1 引　　言

众多研究结果表明，天然蛋白质结构对于氨基酸的替换具有非常显著的“耐受性”。因此很多不同的氨基酸序列可以编码产生给定的三维结构[1~7]。

我们利用这一“耐受性”来发展蛋白质设计的总体策略。这一称为“二元码”的策略是基于极性与非极性氨基酸适当的组合能指导多肽链折叠成相应的二级结构元素，同时使得包埋的非极性氨基酸形成所需的三级结构[8~10]。设计的“二元组图”利用了蛋白质二级结构中天然具备的周期性：α螺旋具有每圈3.6个残基的重复周期性，而β股（β-strand）具有交替周期性（图9.1）。因此，设计为双亲性α螺旋的二元组图序列应该在第三或第四位放上非极性氨基酸。相对应，设计一个双亲性β股应该在序列中交替安插极性和非极性氨基酸残基。在“二元码”策略中，蛋白质侧链的精确三维堆积并不需要预先确定。因此，在一个二元组图序列库中，每一个极性和非极性残基的侧链可以变化得非常剧烈，从而产生巨大的组合多样性。

二元组图蛋白质的组合库是由合成基因的组合库表达出来的。每个基因编码一个不同的氨基酸序列，但是在同一个给定组合库中的所有序列具有同样的极性和非极性残基图形。遗传代码（图9.2）的组成分布使得这种序列简并性成为可能。简并密码NTN编码非极性氨基酸残基，而简并密码NAN编码极性氨基酸残基（N表示A、G、T、C的混合物；见9.2.2节中相关密码子的利用）。利用这些简并密码子，非极性位置上的氨基酸可以是苯丙氨酸（Phe）、亮氨酸（Leu）、异亮氨酸（Ile）、甲硫氨酸（Met）或者缬氨酸（Val）；而极性位置的氨基酸则是谷氨酸（Glu）、谷氨酰胺（Gln）、天冬氨酸（Asp）、天冬酰胺（Asn）、赖氨酸（Lys）或者组氨酸（His）（图

9.2)。避免终止密码子的策略在 9.2.2.1 节中叙述。

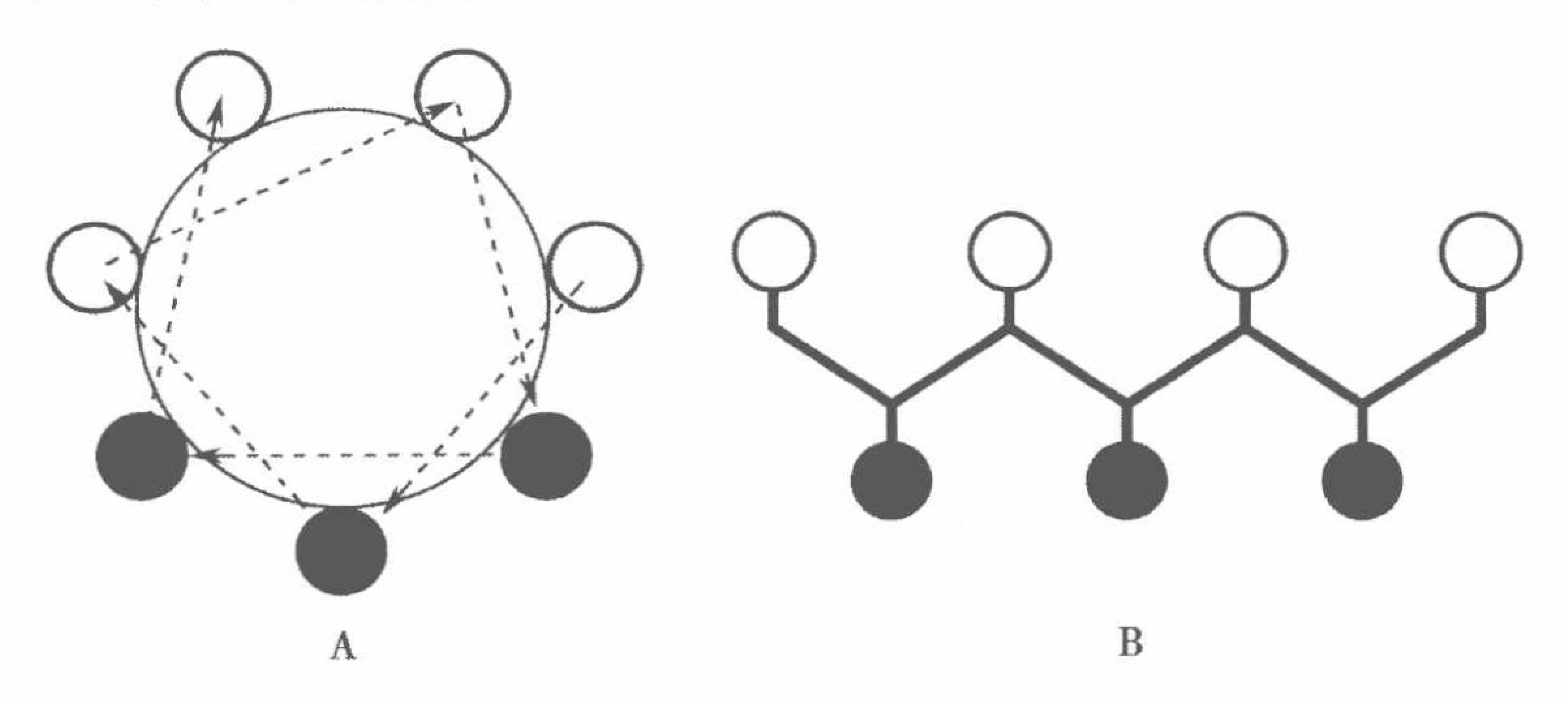


图 9.1 利用 α 螺旋和 β 股二级结构中天然具备的周期性，设计极性（○）与非极性氨基酸残基（●）的二元组图产生相应的二级结构：A. α 螺旋具有每圈 3.6 个残基的重复周期性。因此通过在第三或第四位放上非极性氨基酸可设计出双亲性 α 螺旋，一侧为极性，另一侧为非极性。此图画出了 7 个氨基酸残基的位置。如果要设计更长的 α 螺旋，二元组图的设计要满足每圈 3.6 个残基的周期性来维持整段螺旋的双亲性。B. β 链具有极性和非极性氨基酸残基的交替周期性。这种组图使得 β 链的一侧为极性而另一侧为非极性。

开头碱基 \ 中间碱基	T	C	A	G
T	TTT Phe	TCT Ser	TAT Tyr	TGT Cys
	TTC Phe	TCC Ser	TAC Tyr	TGC Cys
	TTA Leu	TCA Ser	TAA Stop	TGA Stop
	TTG Leu	TCG Ser	TAG Stop	TGG Trp
C	CTT Leu	CCT Pro	CAT His	CGT Arg
	CTC Leu	CCC Pro	CAC His	CGC Arg
	CTA Leu	CCA Pro	CAA Gln	CGA Arg
	CTG Leu	CCG Pro	CAG Gln	CGG Arg
A	ATT Ile	ACT Thr	AAT Asn	AGT Ser
	ATC Ile	ACC Thr	AAC Asn	AGC Ser
	ATA Ile	ACA Thr	AAA Lys	AGA Arg
	ATG Met	ACG Thr	AAG Lys	AGG Arg
G	GTT Val	GCT Ala	GAT Asp	GGT Gly
	GTC Val	GCC Ala	GAC Asp	GGC Gly
	GTA Val	GCA Ala	GAA Glu	GGA Gly
	GTG Val	GCG Ala	GAG Glu	GGG Gly

图 9.2 遗传代码的组成分布使得序列简并性成为可能，通过确定密码子的中间位置碱基，设计合成基因的组合库，表达极性与非极性残基的二元组图蛋白质组合库。对于非极性氨基酸残基位置，简并密码 NTN（N 表示 A、G、T、C 的混合物）可以编码苯丙氨酸（Phe）、亮氨酸（Leu）、异亮氨酸（Ile）、甲硫氨酸（Met）或者缬氨酸（Val）。对于需要极性残基的位置，简并密码 NAN 可编码组氨酸（His）、谷氨酰胺（Gln）、天冬酰胺（Asn）、赖氨酸（Lys）、天冬氨酸（Asp）或者谷氨酸（Glu）。对于 NAN 密码子，为避免终止密码子，碱基 T 不能出现在初始位置。

本章概述了应用二元组图设计新型蛋白质库的方法。利用我们实验室的具体实例，主要集中在设计全 α 螺旋蛋白质和全 β 片层蛋白质上。对于设计蛋白质组合库的综述，参见参考文献［11］、［12］。

9.2　材料与方法

9.2.1　设计结构模板

二元组图可以应用于一个蛋白质中任何双亲性α螺旋或β链部分。尽管我们实验室集中于设计全新蛋白质，二元码策略也可以用于已知蛋白质的局部区域，如活性中心，蛋白质核心的一部分，或者接触界面[13]。对于全新蛋白质的设计，这个策略能否成功取决于蛋白质结构模板设计的好坏。下面我们介绍设计结构模板需要考虑的几个因素。

9.2.1.1　二元组图的区域

1）α螺旋设计

二元组图利用蛋白质二级结构中固有的周期性。α螺旋具有每圈3.6个残基的重复周期性（图9.1A）。要设计一节双亲性α螺旋二级结构片段，可以用二元图形P-N-P-P-N-N-P（P表示极性残基，N表示非极性残基）。我们最初的α螺旋设计集中在四螺旋束模体上（图9.3）。在这个结构中，每个螺旋疏水面都朝向束的中间核心，而各螺旋的亲水面则暴露在水溶液环境中。P-N-P-P-N-N-P图样有利于形成双亲性α螺旋二级结构，使得形成所需三级结构之后所有的非极性氨基酸侧链都被包埋在内部。从我们设计的四螺旋束蛋白质库中，50多个蛋白质被纯化及表征。所有都具有典型的α螺旋圆二色（CD）光谱（见注1）。更值得一提的是，此研究中有好几个蛋白质都表现出天然蛋白的性质。例如，核磁共振化学位移色散，协同性化学和热变性，以及慢速氢/氘交换率等[14~18]。

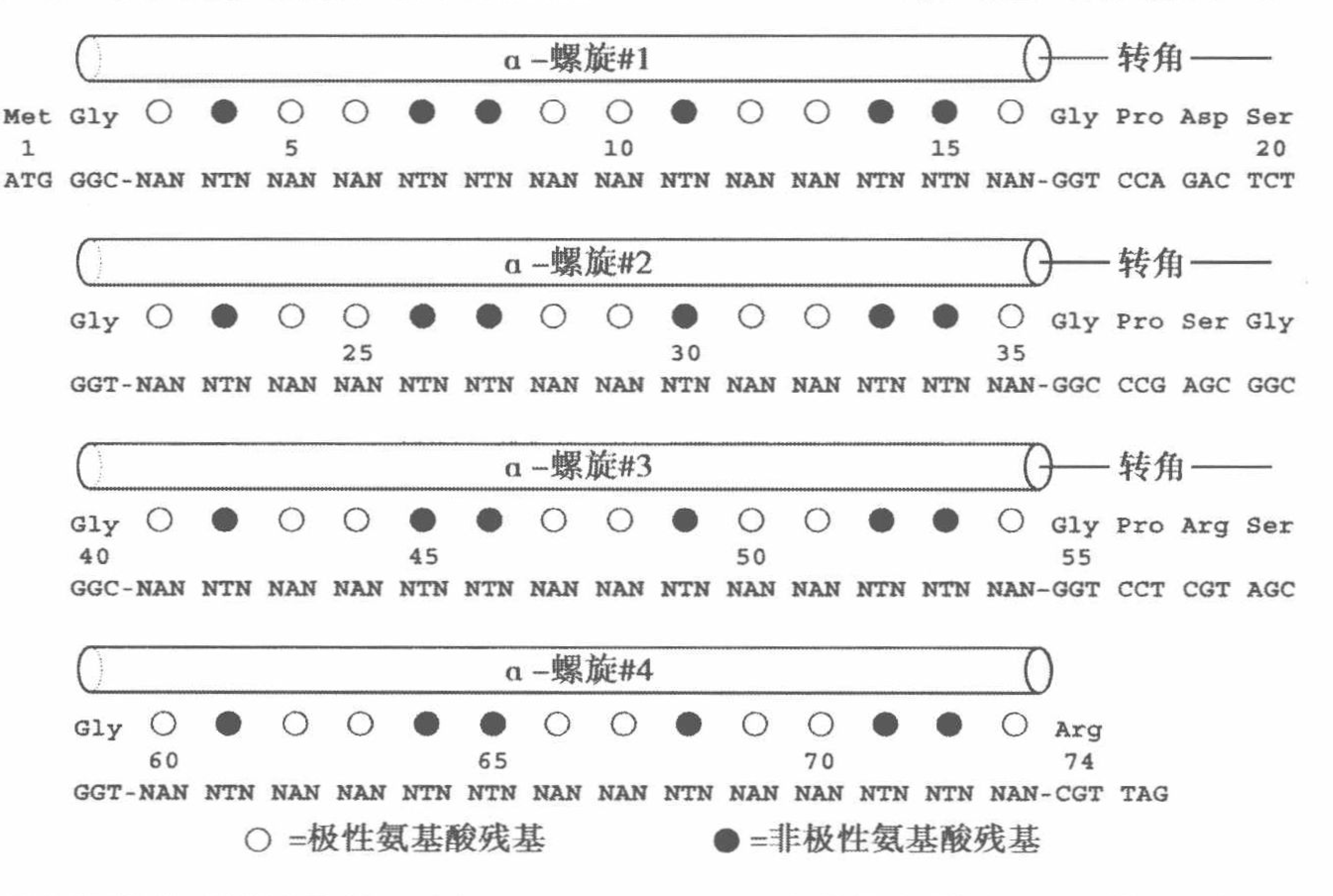


图9.3　最初四螺旋束设计模板。非极性残基位置（●）由简并密码子NTN编码，极性残基位置（○）由简并密码子NAN编码。设计中确定的残基位于N端和C端以及螺旋之间的转角位置。

2）β片层设计

双亲性β链具有…P-N-P-N…的交替周期性（图9.1B）。基于这种周期性，合成基因的组合库可以编码产生β片层结构蛋白质。极性残基组成β片层的一面，而非极性残

基组成相反的另一面。我们最初设计了具有6个β链的片层结构，每一股都具有二元组图 P-N-P-N-P-N-P[9]。利用合成基因克隆到大肠杆菌中表达该库中的蛋白质，研究中所有的蛋白质都形成了β片层二级结构，具有典型的β片层圆二色光谱，217nm 处有一个低谷（见注1）。从这个最初的组合库中得到的β片层蛋白质可以自组装成为淀粉样纤维沉淀[9]。这些纤维将非极性残基包埋在疏水核心，而极性残基则暴露在溶液中。

如果把上述的β片层蛋白质置于具有极性/非极性界面的非均一环境中，它们会形成不同的结构。例如，在空气/水界面，这些蛋白质自组装成平整的β片层单层，非极性残基向上指向空气，而极性残基向下指向水面[19]。或者，在水和高度有序的热熔石墨非极性界面，二元组图的β片层序列在石墨表面经历模板指导的组装进而产生了高度有序的结构[20]。

9.2.1.2 固定区域

实验中，经常需要固定蛋白质的部分序列（即没有组合差异），特别是当目标序列很长的时候。设计合成基因库时，这部分固定的区域可以作为单链合成寡核苷酸退火时相互识别的位点，并引导互补链的酶合成（图9.4；由单链寡核苷酸进行全长基因组装部分见9.2.3）。

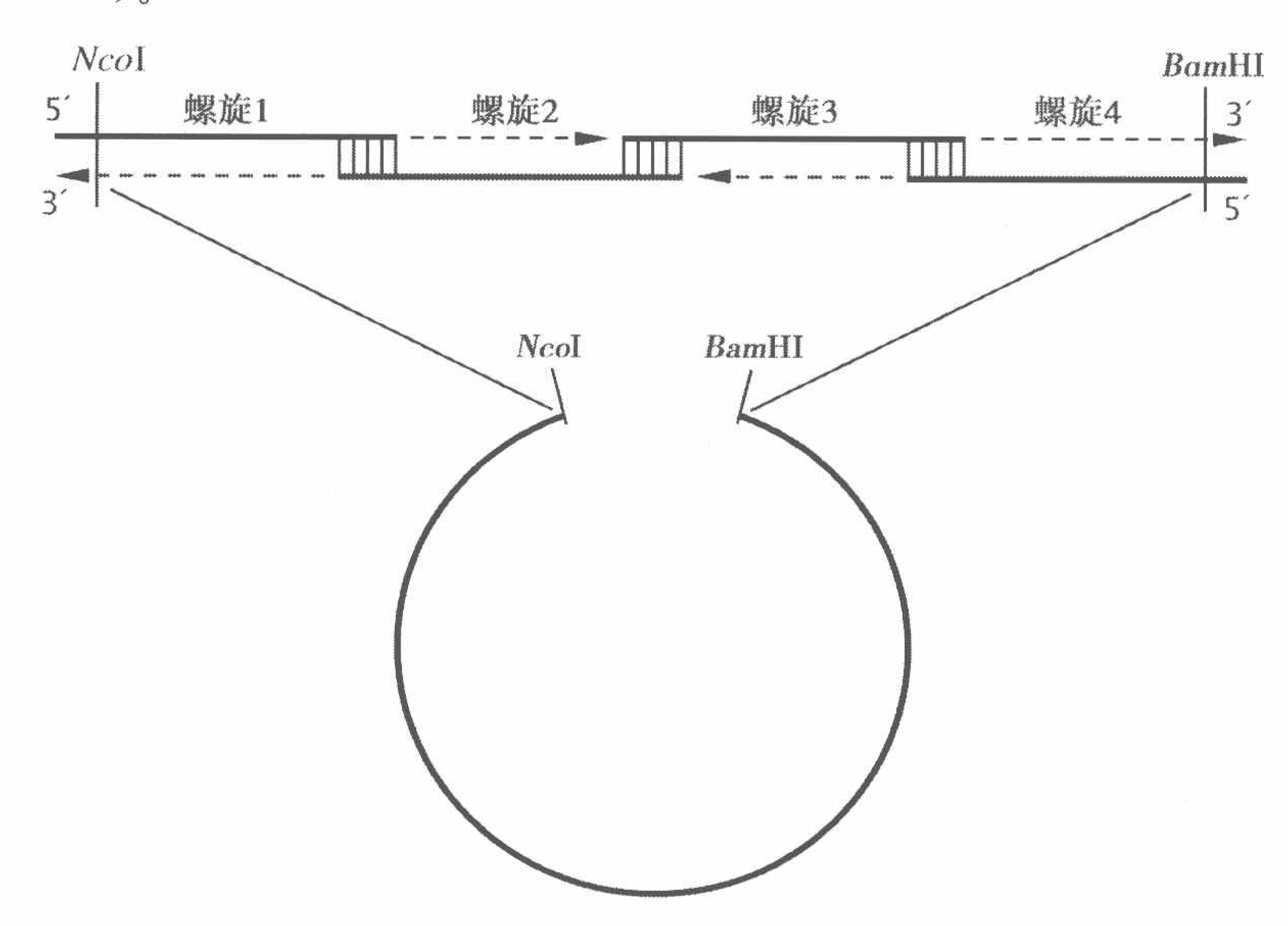


图9.4 应用4段单链寡核苷酸组装全长基因。单链寡核苷酸5′端和3′端的固定区域用作退火及启动酶合成互补链（利用DNA聚合酶）的位点。这样得到的全长基因可以插入表达载体中。

较短的或中等长度的单链寡核苷酸通常用于编码单独二级结构单元的二元组图。这些寡核苷酸3′端和5′端的非简并固定区域通常用于编码固定的转角区域（图9.3和图9.4；参考文献［8］和［9］）。位于这些转角区域的氨基酸残基的选择则基于下面的统计及推理准则。

（1）转角区域氨基酸残基的选择基于位置偏好的原则。例如，在初始的四螺旋束库中，螺旋两端的“N帽子”（N-cap）和“C帽子”（C-cap）区域放置甘氨酸残基（图9.3；参考文献［8］）。天然蛋白中这些位置也经常出现甘氨酸残基[21]。在C帽子后面

的位置，加入脯氨酸残基来终止螺旋结构。但某些情况下，脯氨酸残基可能并不适合，因为其引起的肽键顺反异构可能造成蛋白质的多构象（非单一构象）。在β片层库中，转角区域的设计则基于已知结构的天然蛋白质中氨基酸残基组合的“转角可能性”（见注 2；参考文献［22］）。

（2）转角区域的序列中可引入限制性位点。这在基因扩增中常常很有用（见 9.2.3；参考文献［8］）。

（3）固定区域的长度应满足序列特异性退火的需要。重叠 10～15 个碱基的寡核苷酸对通常用于退火反应。亦可通过固定转角区域前后密码子中的 1 或 2 个碱基来帮助退火。例如，合成寡核苷酸链（5′-NAN-NTN-NTN-NAN-GGT-CCT-CGT-AGC-3′）画线部分的 12 个碱基为固定的部分，编码四残基转角区域。其之前为编码极性残基的 NAN，如果将第三个碱基固定，如 G，密码子变为 NAG，则在序则特异性退火时就产生额外的两个固定碱基（5′-NAN-NTN-NTN-NAG-GGT-CCT-CGT-AGC-3′）。同时，因只固定了密码子的第 2、第 3 个碱基（第 1 个碱基可变），氨基酸的多样性被保留。

除了转角区域外，N 端和 C 端的序列也需要固定。这些区域的固定序列通常是将基因克隆到表达载体时所需的。末端设计的一些准则如下。

（1）N 端放置起始甲硫氨酸，这是胞内重组表达所需要的。

（2）可在固定区域安插含有芳香族生色团的残基（如酪氨酸、苯丙氨酸），这样有利于蛋白质的纯化及浓度的测定[9,18]。这些芳香族的氨基酸残基可放在固定转角区域或是链的一侧末端。在我们设计的一些库中，将一个酪氨酸残基安插在起始甲硫氨酸后面，这样不但可以引入生色团，也可防止体内实验中甲硫氨酸被切除[23～26]。

（3）设计蛋白质的 C 端残基应为带电荷的极性残基。因为蛋白质的 C 端序列可能会影响其胞内降解的速率，带电荷残基的加入会延长其胞内的半衰期[27～29]。因此，在四螺旋库中，就设计了半胱氨酸残基位于 C 端[8,18]。此外，如果 α 螺旋的 C 端为带正电荷的残基（N 端为带负电荷的残基），则可以通过与螺旋的偶极作用稳定蛋白质[30]。

9.2.1.3　设计三级结构的考虑

所设计的模板最终能否成功取决于其编码的蛋白质的性质。如果目标是产生良好折叠的球蛋白，那么设计的模板就不但要求有很好的二级结构，也需要有三级结构。设计的模板要足够长以形成折叠良好的三维结构，同时又要尽量短使得正确扩增大的基因文库得以实现。我们设计的第一代四螺旋束（包含 74 个残基）库中的许多蛋白质都形成了类似溶融球结构的中间物[8,14～17]。为了研究二元码策略编码类似天然蛋白质三级结构的可能，我们设计了第二代二元组图蛋白质库[18]。这个库是以早先设计的 74 残基库中的序列蛋白质 86 为基础进行改造设计而来的。对蛋白质 86 的主要改动是给每个四螺旋加上 6 个不同组合的氨基酸残基，使得设计的第二代蛋白质与天然四螺旋蛋白质大小相似。增加的 24 个残基的设计同样遵循二元组图法则。

从第二代库中随机抽选 5 个序列进行性质鉴定，发现它们全部都比母蛋白 86 更加稳定[17]。同时，它们的核磁共振（nuclear magnetic resonance，NMR）波谱谱线大部分为分散的，显示很好解析的核 Overhauser 效应（nuclear overhauser effect，NOE）交叉峰，表示得到了独特的、良好折叠的三级结构[18]。其中的两个蛋白质的结构通过 NMR 获得解析，均是所设计的有序的四螺旋束结构[12,31]。

9.2.2 密码子的应用

正如 9.1 节中提到的，简并密码子 NAN 和 NTN 分别编码极性和非极性氨基酸残基（图 9.2）。然而，在 N 位置仅简单的使用等量的 A、C、T、G 碱基的混合物会给序列引入不良特性。更为重要的是，未加限制的 NAN 密码子有很高的概率编码终止密码子（即每 16 个 NAN 密码子中就有 2 个终止密码子，概率为 12.5%）。下面是我们使用的设计极性与非极性密码子的一些准则。

9.2.2.1 NAN（极性）密码子

（1）NAN 密码子的第一个碱基使用等比例的 G、C 和 A 的混合物，T 不使用，这样避免了终止密码子和酪氨酸残基的引入（见注 3）。如果 4 种碱基都加入，则 T 碱基会形成终止密码子 TAG 和 TAA。

（2）NAN 的第三个碱基的设计可以优先选择一些碱基。如果使用等比例的 G、C、A、T 碱基的混合物，则产生组氨酸、谷氨酰胺、天冬酰胺、赖氨酸、天冬氨酸及谷氨酸的概率是相同的。但 NAN 编码的某些残基有形成 α 螺旋的固有倾向[32~35]。因此通过在第 3 个碱基的位置排除 T 碱基，则更多的形成谷氨酰胺、赖氨酸和谷氨酸，而不是组氨酸、天冬酰胺和天冬氨酸。这样就正好符合这些残基形成 α 螺旋的偏好[32~35]。

9.2.2.2 NTN（非极性）密码子

如果在 NTN 密码子的第 1 个和第 3 个 N 位置都放置等比例的 4 种碱基，那么编码亮氨酸的比率是甲硫氨酸的 6 倍（即 6 个亮氨酸密码子对 1 个甲硫氨酸密码子）。并且这种等比例的设计会造成内部 1/4 的疏水残基都将是缬氨酸，缬氨酸不易形成 α 螺旋，因此可能不适于某些设计。通过改变 N 位置上 4 种碱基的比率，各种疏水残基的相对含量会发生变化。例如，在初始的四螺旋束库，NTN 密码子的第 1 个碱基位置上 A∶T∶C∶G 的摩尔比为 3∶3∶3∶1，而第 3 个位置上只包含等摩尔的 G 和 C 碱基[8]。通过这种混合物的偏好，缬氨酸在疏水残基中所占的比率下降到 10%，而亮氨酸残基出现的频率也下降为甲硫氨酸的 3 倍。

9.2.2.3 宿主表达系统中密码子的应用

不论是固定区域或是组合区域，设计库中的 DNA 序列都应优先使用宿主表达系统喜好的密码子。例如，设计简并密码子的第三个碱基仅包含 G 和 C（而非 4 种碱基）以满足大肠杆菌对密码子的偏好性。对于宿主细胞很少使用的密码子，如大肠杆菌表达系统中 CGA、AGA 和 AGG（编码精氨酸），CTA（编码亮氨酸），CCC（编码脯氨酸），ATA（编码异亮氨酸）都应尽量避免，因为包含稀有密码子的基因很难被表达[37]。其他的（非大肠杆菌）表达系统也都有各自的密码子偏好性，在设计中应该考虑。

9.2.3 全长基因的组装

通常我们使用短的单链寡核苷酸组装全长基因（图 9.4）。这样可以减少直接合成长链核苷酸引起的固有误差（多数由于缺失或移码引起）。5′端和 3′端的固定区域是单链寡核苷酸退火杂交并引导酶（DNA 聚合酶）合成互补链的位点。应用这一策略可以

获得更多的正确编码新型目标蛋白的“无误基因”。

在合成半随机寡核苷酸时，一些是编码链（有义链），另一些是非编码链（无义链）。通常每条寡核苷酸链都编码一个独立的二元组图二级结构片段。用这样的片段组装全长基因使得单独的α螺旋和β股可作为独立的模块被设计和使用，因此增加了二元码策略的多样性（见注 4）。

在我们最初设计的四螺旋束库中，使用 4 条合成的寡核苷酸链构建全长基因。每条链都编码一个单独的螺旋及转角。如 9.2.1.2 中所述，转角区域都是固定的（即没有简并），以作为 DNA 聚合酶合成互补链时的引导位点（图 9.4；参考文献 [8]）。

我们尝试了多种方法进行基因组装。比如同时构建了两个半基因库，将两者连接在一起生成编码全长蛋白质的基因库[8]。为了确保做到正确的头尾相连，可在连接部位的固定区域设计非回文结构的限制性酶切位点[8]。其他的一些组装全长基因的方法包括多种聚合酶链反应策略（如重叠延伸 PCR），这些方法在我们构建的多个文库中都曾使用。

9.3　注

1. α螺旋和β片层库均由同样的二元码氨基酸残基组成。因此，并不是氨基酸组成的不同赋予所设计的蛋白质不同特征[8~10]。同样也不是因为序列长短的不同造成蛋白质的不同特性。不需考虑长度，周期序列 P-N-P-P-N-N-P 总是形成α螺旋二级结构，而 P-N-P-N-P-N-P 周期序列则总是形成β片层二级结构（在相同的实验条件下检测）。二元组图本身就是这两种不同结构的序列库的不同之处。

2. 对于二元码密码子 NAN 和 NTN，可编码 6 种极性氨基酸残基（谷氨酸、天冬氨酸、赖氨酸、天冬酰胺、谷氨酰胺及组氨酸）和 5 种非极性氨基酸残基（缬氨酸、甲硫氨酸、异亮氨酸、亮氨酸和苯丙氨酸）。除了这 11 种可变残基之外，可在序列的固定区域加入许多其他残基。例如，我们新近设计的包含 102 个残基的四螺旋束库就包含了 20 种氨基酸中的 17 种[18]。只去除了丙氨酸、脯氨酸和半胱氨酸。因为在天然蛋白质中，丙氨酸在蛋白质的表面和核心都有出现，其在二元码中的极性或非极性性质比较模糊。而脯氨酸是个特例，因其 Φ 角受限制，仅能在结构的某些固定区域使用。半胱氨酸仅在需要设计二硫键或金属结合位点时才使用。

3. 通过在极性密码子 NAN 的第一个碱基位置不使用 T 碱基，避免酪氨酸的出现。这是因为天然蛋白质中酪氨酸并不是完全的极性残基，经常会出现在疏水内核的位置。因此只有极性较强的残基（组氨酸、谷氨酰胺、天冬酰胺、赖氨酸、天冬氨酸和谷氨酸）才设计在蛋白质表面位置。

4. 用聚丙烯酰胺凝胶电泳法纯化合成的寡核苷酸非常必要。这样可减少将截短的寡核苷酸引入库中的可能性。虽然这一纯化步骤减少了 DNA 的含量（以及潜在的多样性），但文库中基因的质量得到了显著提高。

致谢　本项目得到下列基金支持：国家卫生研究院（NIH）（R01-GM62869）、陆军科研办公室（DAAG55-98-1-0084）、自然科学基金（NSF）（MRSEC DMR98-09483）以及国防高级研究计划（n00173-01-1-0015）。

参考文献

1. Lim, W. A. and Sauer, R. T. (1989) Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* 339, 31-36.
2. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* 247, 1306-1310.
3. Axe, D. D., Foster, N. W., and Fersht, A. R. (1996) Active barnase variants with completely random hydrophobic cores. *Proc. Natl. Acad. Sci. USA* 93, 5590-5594.
4. Gassner, N. C., Baase, W. A., and Matthews, B. W. (1996) A test of the “jigsaw puzzle” model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc. Natl. Acad. Sci. USA* 93, 12155-12158.
5. Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., et al. (1997) Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* 4, 805-809.
6. Silverman, J. A., Balakrishnan, R., and Harbury, P. B. (2001) Reverse engineering the $(\beta/\alpha)_8$ barrel fold. *Proc. Natl. Acad. Sci. USA* 98, 3092-3097.
7. Lau, K. F. and Dill, K. A. (1990) Theory for protein mutability and biogenesis. *Proc. Natl. Acad. Sci. USA* 87, 638--642.
8. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262, 1680-1685.
9. West, M. W., Wang, W., Patterson, J., Mancias, J. D., Beasley, J. R., and Hecht, M. H. (1999) *De novo* amyloid proteins from designed combinatorial libraries. *Proc. Natl. Acad. Sci. USA* 96, 11211-11216.
10. Xiong, H., Buckwalter, B. L., Shieh, H. M., and Hecht, M. H. (1995) Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proc. Natl. Acad. Sci. USA* 92, 6349-6353.
11. Moffet, D. A. and Hecht, M. H. (2001) *De novo* proteins from combinatorial libraries. *Chem. Rev.* 101, 3191-3203.
12. Hecht, M. H., Das, A., Go, A., Bradley, L. H., and Wei, Y. (2004) *De novo* proteins from designed combinatorial libraries. *Protein Sci.* 13, 1711-1723.
13. Taylor, S. V., Walter, K. U., Kast, P., and Hilvert, D. (2001) Searching sequence space for protein catalysts. *Proc. Natl. Acad. Sci. USA* 98, 10596-10601.
14. Roy, S., Ratnaswamy, G., Boice, J. A., Fairman, F., McLendon, G., and Hecht, M. H. (1997) A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. *J. Am. Chem. Soc.* 119, 5302-5306.
15. Roy, S., Helmer, K. J., and Hecht, M. H. (1997) Detecting native-like properties in combinatorial libraries of *de novo* proteins. *Folding Des.* 2, 89-92.
16. Roy, S. and Hecht, M. H. (2000) Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. *Biochemistry* 39, 4603-4607.
17. Rosenbaum, D. M., Roy, S., and Hecht, M. H. (1999) Screening combinatorial libraries of *de novo* proteins by hydrogen-deuterium exchange and electrospray mass spectrometry. *J. Am. Chem. Soc.* 121, 9509-9513.
18. Wei, Y., Liu, T. I. P., Sazinsky, S. L., Moffet, D. A., and Hecht, M. H. (2003) Well folded *de novo* proteins from a designed combinatorial library. *Protein Sci.* 12, 92-102.
19. Xu, G., Wang, W., Groves, J. T., and Hecht, M. H. (2001) Self-assembled monolayers from a designed combinatorial library of *de novo* β-sheet proteins. *Proc. Natl. Acad. Sci. USA* 98, 3652-3657.
20. Brown, C. L., Aksay, I. A., Saville, D. A., and Hecht, M. H. (2002) Template-directed assembly of a *de novo* designed protein. *J. Am. Chem. Soc.* 124, 6846-6848.
21. Richardson, J. S. and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices. *Science* 240, 1648-1652.

22. Hutchinson, E. G. and Thornton, J. M. (1994) A revised set of potentials for β-turn formation in proteins. *Protein Sci.* 3, 2207-2216.

23. Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G., and Blanquet, S. (1989) Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. USA* 86, 8247-8251.

24. Dalboge, H., Bayne, S., and Pedersen, J. (1990) *In vivo* processing of N-terminal methionine in *E. coli*. *FEBS Lett.* 266, 1-3.

25. Tsunasawa, S., Stewart, J. W., and Sherman, F. (1985) Amino-terminal processing of mutant forms of yeast iso-1-cytochrome c. The specificities of methionine aminopeptidase and acetyltransferase. *J. Biol. Chem.* 260, 5382-5391.

26. Huang, S., Elliott, R. C., Liu, P. S., et al. (1987) Specificity of cotranslational amino-terminal processing of proteins in yeast. *Biochemistry* 26, 8242-8246.

27. Bowie, J. U. and Sauer, R. T. (1989) Identification of C-terminal extensions that protect proteins from intracellular proteolysis. *J. Biol. Chem.* 264, 7596-7602.

28. Parsell, D. A., Silber, K. R., and Sauer, R. T. (1990) Carboxy-terminal determinants of intracellular protein degradation. *Genes Dev.* 4, 277-286.

29. Milla, M. E., Brown, B. M., and Sauer, R. T. (1993) P22 Arc repressor: enhanced expression of unstable mutants by addition of polar C-terminal sequences. *Protein Sci.* 2, 2198-2205.

30. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., and Baldwin, R. L. (1987) Tests of the helix dipole model for stabilization of alpha-helices. *Nature* 326, 563-567.

31. Wei, Y., Kim, S., Fela, D., and Hecht, M. H. (2003) Solution structure of a de novo protein from a designed combinatorial library. *Proc. Natl. Acad. Sci. USA* 100, 13270-13273.

32. Chou, P. Y. and Fasman, G. D. (1978) Empirical predictions of protein conformation. *Annu. Rev. Biochem.* 47, 251-276.

33. Fasman, G. D. (1989) *Prediction of Protein Structure and the Principles of Protein Conformation*. Plenum, New York, NY.

34. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*. 2nd ed., Freeman, New York, NY.

35. Pace, C. N. and Scholtz, J. M. (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* 75, 422-427.

36. DeBoer, H. A. and Kastelein, R. A. (1986) in *Maximizing Gene Expression* (Reznikoff, W. and Gold, L., eds.), Butterworth, Stoneham, MA, pp. 225-285.

37. Kane, J. F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 6, 494-500.

10　利用核苷酸交换和剪切技术进行 DNA 碎裂和定向进化

Sabine C. Stebel，Katja M. Arndt，
and Kristian M. Müller

概要　利用 DNA 混编模仿自然进化过程是优化 DNA 和蛋白质性质的常用方法。这里我们介绍这类方法的一个新进展，即利用标准的聚合酶链反应（PCR）放大基因文库时与其他 4 种标准 dNTP 一起掺入 dUTP 作为确定 DNA 碎裂位点的交换核苷酸。掺入的尿嘧啶碱基可用尿嘧啶-DNA-糖基化酶切除，随后 DNA 主链用哌啶切开。这个寡聚核苷酸库的重组装由内引物延伸步骤及高保真聚合酶来增加产率，最后由 PCR 扩增。变性聚丙烯酰胺尿素凝胶电泳表明这个方法可以产生大小可调的 DNA 片段，大小范围取决于 dUTP∶dTTP 比率。对于一个模式蛋白，氯霉素乙酰转移酶 I（CAT），利用包含 33%dUTP 的 PCR 产生的混编基因库测序显示了大约 0.1%的低突变率，并且，在没有进行片段分离的情况下，平均亲本片段大小为 86 个碱基。因此，核苷酸交换和剪切技术（NExT）DNA 混编可重复性好并且容易实行，使之优于相应同类技术。另外，NExT 碎裂的结果可以利用计算机软件，NexTProg 预测出来。

关键词　DNA 混编；定向进化；体外进化；蛋白质工程；UDG；dUTP；氯霉素乙酰基转移酶（CAT）

10.1　简　　介

近年来，化学、制药和食品加工等行业对“特性酶”（tailor-made enzyme）的需求越来越多。因为关于蛋白质的折叠、相互作用及如何稳定都并未完全研究清楚，到现在设计特性蛋白的途径仍然很复杂，并且要求耗时的多轮的设计和检验。采用重复进行突变、筛选和混编的进化步骤是获得目标特性的快速有效的手段，即使蛋白质结构特征并不知晓。

多种方法可用于基因库的体外重组[1]。然而使用起来无一不是有缺陷或是有困难的。例如，发展的非常成熟的 Stemmer 方法，通过优化酶切条件，使用 DNA 酶切割 DNA 获取所需大小的片段[2,3]。为进一步筛选大小合适的片段，通过琼脂糖凝胶电泳纯化酶切片段，再经内部引物延伸反应再组装，最后使用常规 PCR 扩增。然而，精确控制 DNA 酶的酶切条件，如使用的酶及 DNA 的量、反应时间及温度等非常复杂，需要精细的优化摸索。其他的一些方法，如交错延伸程序（staggered extension process，StEP[4]）和随机引物法（random priming）[5]都受 DNA 组成的限制，并且由于缺少对重组及产生片段大小范围的控制而更为复杂难辨。因此欲使用上述的方法得到理想的结果都需要非常缜密的优化筛选。

不同于上述的方法，我们设计了一种高效、稳定的新方法。核苷酸交换和剪切技术（NExT）进行 DNA 混编是以“交换核苷酸”的随机引入为基础。这些交换核苷酸在 DNA 序列中的出现和位置就指明了其产生的裂解部位，不需要进一步的调整。

应用 NExT 方法提高氯霉素乙酰转移酶 I（CAT，可产生对氯霉素的抗性）截短突变体功能的研究对 NExT 方法进行了很好的发展和检验。在基因水平上对四组截短突变体独立开展定向进化实验。第一组 CAT 突变体库为 N 端截短 10 个氨基酸残基（CAT _ Nd10），第二组突变体库为 C 端截短 9 个氨基酸残基（CAT _ Cd9），第三组为 C 端截短 26 个氨基酸残基（CAT _ Cd26），第四组为 N 端和 C 端分别截短 10 个和 9 个氨基酸残基（CAT _ Nd10 _ Cd9）。从易错 PCR（error-prone PCR）多样化步骤中挑选 3～6 个产物作为检测库可以获得 NExT DNA 混编的详细数据[6]。

CAT _ Nd10 可以在含有 25μg/ml 氯霉素的平板上生长，而 CAT _ Cd9 则根本不生长。应用基于易错 PCR 和 NExT DNA 混编的完整定向进化策略提高二者的酶活性。通过优化，二个库均有若干克隆可在含有 400μg/ml 氯霉素的平板上生长（未检测更高的浓度），证明了此种技术的有效性。同时，这种优化的方法也应用于根据我们的计算程序选出的含 30% dUTP 的 TEM-1 型 β-内酰胺酶（见 10.3.9）。这种裂解和再组装在实验的第一步就引入，只需确保足够的起始材料，节省了全部的前导和中间检测以及分析步骤（数据略去）。

本章概述了优化的 NExT 混编方法的全部步骤，从克隆开始（见 10.3.1），到尿苷酸交换 PCR（见 10.3.2），酶解反应及化学剪切（见 10.3.3），片段纯化（见 10.3.5），及最终的再组装和扩增（见 10.3.7）。对各步骤可调整的部分也分别进行了讨论。同时，给出了关于混编过程详细特性的分析方法（见 10.3.4 和 10.3.6），比较了不同方法的交换率、平均片段长度和变异率（见 10.3.8）。最后，介绍了一种预测 NExT 断裂参数和结果的计算程序（见 10.3.9）。

NExT 方法基于预先确定的合理的 dUTP∶dTTP 的比率，操作简单，因此非常适于短的基因片段和大的基因组合的混编。由于 NExT 混编方法的稳定和简便性，就算之前很少涉及这个领域的工作者也能应用此方法解决 DNA 和蛋白质水平上的生物化学领域的问题。

10.2　材　　料

10.2.1　克隆

（1）切割目标基因和质粒的合适的限制性内切核酸酶。这里使用 *Xba*I 和 *Hind*III。

（2）T4 DNA 连接酶（不同厂家）。

（3）感受态细胞（大肠杆菌 RV308 和 XL-1）。

（4）质粒载体 pLisc-SAFH11[7]。

10.2.2 尿苷酸交换 PCR

（1）5′和 3′引物各 1 条，满足基因扩增需要，并且包含克隆位点。

（2）标准 *Taq* 聚合酶（不同厂家）。

（3）10×PCR 缓冲液（不同厂家），如 160mmol/L 硫酸铵，670mmol/L Tris-氯化氢（25℃时 pH 8.8），15mmol/L 氯化镁和 0.1%的吐温-20；或者 100mmol/L Tris 氯化氢（25℃时 pH 9.0），500mmol/L 氯化钠，15mmol/L 氯化镁和 1%的 Triton X-100。

（4）混合 dATP、dTTP、dGTP、dCTP 和 dUTP 的 dNTP 储液（10mmol/L 或 100mmol/L；不同厂家）。

（5）灭菌去离子水。

（6）10×Tris-硼酸-EDTA 缓冲液（TBE 缓冲液）：包含 1mol/L Tris、0.83mol/L 硼酸和 1mmol/L EDTA。

（7）溴化乙锭（EB）溶液：5mg/ml EB 水溶液。

（8）1%的琼脂糖凝胶：1%（*m*/*V*）的琼脂糖溶解在 0.5×TBE 缓冲液中。加热至琼脂糖完全溶解。倒胶前加入 EB（1∶10 000 稀释）。在 0.5×TBE 缓冲液中电泳。

（9）胶回收试剂盒（GE 或 Qiagen 公司）。

（10）PCR 回收试剂盒（GE 或 Qiagen 公司）。

10.2.3 酶解反应和化学切割

（1）大肠杆菌尿嘧啶-DNA-糖基化酶（UDG）（NEB 或 Peqlab biotechnologie GmbH 公司）。

（2）99.9%的哌啶（Sigma 公司）。

10.2.4 尿素变性聚丙烯酰胺凝胶电泳

（1）尿素（纯度＞99.5%）。

（2）30%聚丙烯酰胺，0.8%甲叉双丙烯酰胺（37.5∶1）储液（Carl Roth GmbH 公司）。

（3）10×TBE 缓冲液（见 10.2.2.6）。

（4）10%过硫酸铵（APS）水溶液。

（5）四甲基二乙胺（TEMED）。

（6）去离子甲酰胺。

（7）固定分子质量大小的寡核苷酸，作为分子质量标准或短距离 DNA 标尺，如 100 个碱基梯度标尺（NEB 公司）。

（8）6×溴酚蓝上样缓冲液：70%（*m*/*V*）的蔗糖溶液中加 0.25%（*m*/*V*）的溴酚蓝。

（9）溴化乙锭溶液：5mg/ml EB 水溶液。

10.2.5　片段纯化

10.2.5.1　哌啶切割后的直接纯化

(1) QiaexⅡ试剂盒（Qiagen 公司）。
(2) 乙酸缓冲液：3mol/L 乙酸钠溶液，pH5.3。
(3) 洗脱缓冲液：10mmol/L Tris-氯化氢，pH8.0。

10.2.5.2　通过尿素变性聚丙烯酰胺凝胶电泳纯化

(1) QiaexⅡ纯化：扩散缓冲液：0.5mol/L 乙酸铵，10mmol/L 乙酸镁，1mmol/L EDTA，0.1%十二烷基硫酸钠（SDS），pH8.0。
(2) 提取水溶液：灭菌去离子水。
(3) 乙酸缓冲液：3mol/L 乙酸钠溶液，pH5.3。
(4) 1mol/L 氯化镁。
(5) 异丙醇。
(6) 90%乙醇。
(7) 洗脱缓冲液：10mmol/L Tris-氯化氢，pH8.0。

10.2.6　纯化片段的定量

1∶5000 稀释的 SYBR 绿色荧光染料Ⅱ（分子探针）。

10.2.7　基因重组装和扩增

(1) 10mmol/L dATP、dTTP、dCTP 和 dGTP 的混合物。
(2) Vent DNA 聚合酶（NEB 公司）。
(3) 25mmol/L 硫酸镁。
(4) *Taq* DNA 聚合酶（不同厂家）。

10.3　方　　法

10.3.1　克隆

NExT 混编过程中，使用包含限制性酶切位点的混编引物的配对位点应设置在紧挨着目标基因的两侧，理想的退火温度为 60℃左右。通过 4 加 2 法则估算引物的退火温度：每个 G 碱基和 C 碱基为 4℃，A 碱基和 T 碱基为 2℃。

我们在 NExT 混编过程中使用的基因包括：含 657 个碱基的野生型 CAT 基因（CATwt；SwissProt 编号：P00483；蛋白质数据库（PDB）编号：1NOC：B），编码 N 端 10 个残基截短的突变体，C 端 9 个碱基截短和 N 端、C 端双截短 CAT 突变体（CAT _ Nd10，CAT _Cd9 和 CAT _ Nd10 _ Cd9）以及 C 端 26 个碱基截短突变体

(CAT _ Cd26)。这些基因和全部的混编基因都克隆到质粒载体 pLisc-SAFH1 上[7]，使用 *Xba*I 和 *Hin*d III 双酶切位点替换原始质粒的部分片段。多样性的产生则通过易错 PCR 和尿苷酸交换 PCR（见 10.3.2 节）。扩增野生型和 N 端截短的基因使用 Pr-N-shuffle（5′-ATTTCTAGATAACGAGGGCAA-3′）和 Pr-C-shuffle（5′-ACTTCACAGGTCAAGCTTTC-3′）混编引物；而对于扩增 C 端截短和双截短突变体的基因则使用 Pr-N-shuffle 和 Pr-Cdx-shuffle（5′-CTTCACAGGTCAAGCTTATCA-3′）引物。通过常规方法将这些质粒转化到大肠杆菌感受态细胞中[8]。

10.3.2 尿苷酸交换 PCR

选择尿苷酸作为交换核苷酸，因为许多种 DNA 聚合酶都可将 dUTP 引入基因序列[9]。尿苷酸交换 PCR 在扩增目标基因库的同时，也引入了交换核苷酸，尿苷酸。通过改变 dUTP∶dTTP 的比率可获得理想的 DNA 断裂效果。这一步骤的实现即可通过分析不同 U∶T 比率的实验，详见 10.3.4 节，也可使用我们为此目的开发的程序（见 10.3.9 节）。使用高达 50%比率的 dUTP 并不影响 PCR 产物的得率。只有在完全使用 dUTP 的情况下，PCR 产物的产量大约为其他反应的 1/4（图 10.1A）。

（1）为精确加样量，将 100mmol/L 核苷酸储液用水稀释到 dATP、dGTP 和 dCTP 的终浓度为 10mmol/L，dUTP 和 dTTP 终浓度 1mmol/L。

（2）尿苷酸交换 PCR 混合液包括：50ng 模板（对含 4340 个碱基的质粒为 0.017pmol），引物各 25pmol，dATP、dGTP 和 dCTP 各 0.4mmol/L，0.4mmol/L dUTP∶dTTP 不同比率的混合物，5 U *Taq* DNA 聚合酶，5μl 10×PCR 缓冲液。用水加至终体积 50μl（见注 1 和注 2）。

（3）温度循环程序为 94℃预变性 1min；92℃变性 30s，62℃退火 20s，72℃延伸 2min，共计 25 个循环；最后 72℃再延伸 4min。其中将延伸时间延长至 2min 是因为测试表明这样可以显著的提高产量（数据略去）。根据 PCR 的不同产量，需同时做 4 个或更多的 50μl 反应体系以获得足够的产物（胶回收后应最好约为 7μg DNA；见注 3）。

（4）将得到的 PCR 产物合并，并通过 1%的琼脂糖凝胶电泳分离纯化（见注 4）。如果量足够多的话，在日光下即可从胶上看到浅红色的条带。将条带切割下来后加到一两个胶回收试剂盒的离心柱中，用 50μl 洗脱缓冲液洗脱。

（5）使用波长范围 220～350nm 的分光光度计检测 260nm 的基线校准吸光值，测定 PCR 回收产物的浓度。我们通常以 1∶30 稀释后在 140μl 的比色杯中测量。NExT DNA 混编产物最理想的得率为 7μg DNA。更低的得率可能也行，但足够量的 DNA 可确保下一步的基因重组顺利进行（见 10.3.7）。将所有理想的 DNA 都用于下步反应。

在这项研究中，我们均使用尿苷酸作为交换核苷酸，然而此项技术也同样适用于其他类似物的引入。例如，8-氧化鸟嘌呤（8-oxoguanine）可被 8-氧化鸟嘌呤 DNA 糖基化酶 FPG（foramidopyrimidine-DNA glycosylate）切除[10]。该碱基可用于 AT 富集区，那里 UDG 酶频繁的切割 DNA，也可用于缺少胸腺嘧啶的 GC 富集区。因此，将多种交换核苷酸结合在一起使用可以产生所需大小的片段，用于下一步的再组装。另外，在 PCR 引物中引入交换核苷酸应确保基因库中的这些区域同样被混编。

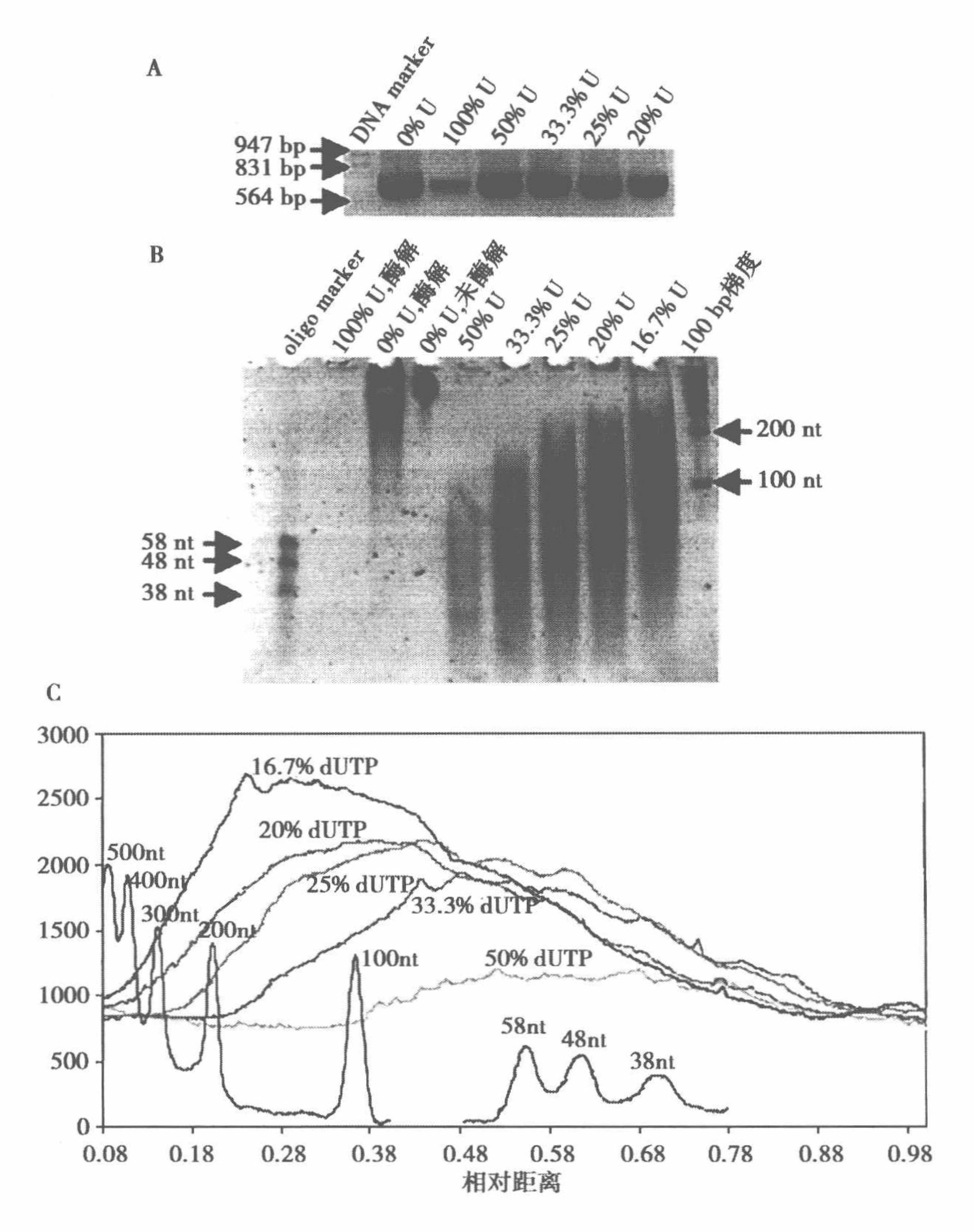


图 10.1　NExT DNA 断裂技术的概述及定量。A. CAT _ Nd10 尿苷酸交换 PCR 产物的 1%琼脂糖凝胶电泳结果：反应中分别使用不同比率的尿嘧啶，计算方法为 U= {c (dUTP) / [c (dUTP)+ c (dTTP)]} × 100。B. 使用 UDG 和哌啶切割 CAT _ Nd10 PCR 产物的尿素聚丙烯酰胺凝胶电泳 EB 染色结果，尝试不同的 dUTP∶dTTP 比率（1∶0、0∶1、1∶1、1∶2、1∶3、1∶4 和 1∶5）确定使用比率的最佳值。条带 1：58bp、48bp 和 36bp 的寡核苷酸混合物作为 DNA Marker；条带 2：使用 100%的 dUTP 进行 PCR 后酶解反应结果；条带 3：使用 0%的 dUTP，酶解；条带 4：使用 0%的 dUTP，未酶解；条带 5：使用 50%的 dUTP，酶解；条带 6：使用 33.3%的 dUTP，酶解；条带 7：使用 25%的 dUTP，酶解；条带 8：使用 20%的 dUTP，酶解；条带 9：使用 16.7%的 dUTP，酶解；条带 10：100bp 碱基梯度 DNA 标尺。注意残留的少量哌啶会使电泳条带轻微扭曲。C. 图 B 中条带 1 及条带 5-10 的密度图，详细描述了使用不同比率 dUTP 得到的片段大小范围。所有 CAT 变体的混编中，我们均使用了 33.3%的 dUTP 比率，产生出的片段长度范围为 30～200bp（粗线所示）。此图先将电泳图经 FluorS Multi-mager 扫描成像后，用 Quantity One (Bio-Rad) 程序计算生成。为使图表清楚，将 100bp 标尺的信号降低了 750 点，将寡核苷酸的信号降低了 500 点。图中“nt”代表寡核苷酸。

10.3.3 酶解反应和化学切割

将尿苷酸交换PCR纯化后的产物进行UDG酶解反应，在交换核苷酸位置切割DNA[11]。该酶对双链和单链DNA均有效，通过对双脱氧尿苷C1′位点的亲核攻击引发水解反应，高特异的去除尿嘧啶基团[12]。哌啶用于经UDG酶切割去除尿嘧啶后的骨架部分[13]。通过高分辨率尿素变性聚丙烯酰胺凝胶电泳分析上述切割反应的结果，其中dUTP的比率从100%到0%不等（图10.1B），进而再用图像分析软件定量（图10.1C）。对于很小的片段，可通过引入放射性标记来更好地观测（见注2）。

（1）UDG酶解反应体系包括45μl PCR纯化产物（含约7μg DNA，见10.3.2节步骤4)，6μl 10× UDG反应缓冲液，2 U大肠杆菌UDG酶。加水至60μl，37℃消化1h（见注6）。

（2）使用哌啶切割DNA反应体系包括终浓度为10%（*m*/*V*）的哌啶（60μl体系中加6.7μl哌啶储液)，在有加热盖子的热循环仪中90℃反应30min（见注7）。

作为采用相对温和的反应条件的尝试，用其他的一些内切核酸酶替代哌啶切割骨架DNA[11]，如内切核酸酶IV[14]，核酸外切酶III[15]，或T4-内切核酸酶V[16]。我们检测了最后一种但否定了其使用。T4内切核酸酶V切割UDG作用后的DNA结果得到的是不完全断裂，这从产物的大小分布可见（图10.2A)。更麻烦的是这一反应必将引起高错误率的出现。应用UDG和T4内切核酸酶V切割CAT野生型基因，胶回收后进行重组装，送其中的6个克隆共计3930bp去测序，结果突变率为1.75%。对于这一结果我们给出了两个原因。首先，T4内切核酸酶V是通过其裂解酶活性切割DNA骨架，其间催化了一个β消除反应，留下了3′端不饱和的醛基（4-羟基-2-戊醛）连在磷酸基团上[11,17]。进一步生成游离磷酸基的化学反应并未完成，如此产生的片段不利于聚合酶的起始。其次，通过片段比较发现，不是所有引入尿苷酸位点的骨架都完成了切割。但这些位置的尿嘧啶都被切掉了，这种碱基缺失的模板会造成错误核苷酸的引入。Miyazaki[18]建议使用大肠杆菌内切核酸酶V，但从提供的数据可见，切割的效率更低，即使延长切割时间（12h）并提高dUTP的比率（75%）仍不能获得短的片段。虽然可以设计进一步的实验解决这些问题，但因哌啶的切割即好又划算，就没有必要再做尝试。

作为哌啶的化学替代品，我们检测了氢氧化钠的效果。将哌啶储液替换为0.5mol/L氢氧化钠溶液，加至DNA体积的10%（*V*/*V*)，90℃反应30min。使用哌啶和氢氧化钠切割相同的尿苷酸交换PCR产物，比较二者变性聚丙烯酰胺凝胶电泳后EB染色的结果。通过图像分析软件，得到的断裂片段的强度分布基本一致。平行进行的重组装反应也得到了相似的结果。这些都证明了哌啶引发的主要反应是碱催化的β消除反应，将两个余下的磷酸基从DNA的糖基上去除下来。因此，如果非常重视安全因素，可用氢氧化钠替代哌啶。然而我们未测序分析氢氧化钠切割片段重组装后的产物是否产生突变。原则上，哌啶更适合引发糖基上的亲核攻击，切割的效果也更为彻底。

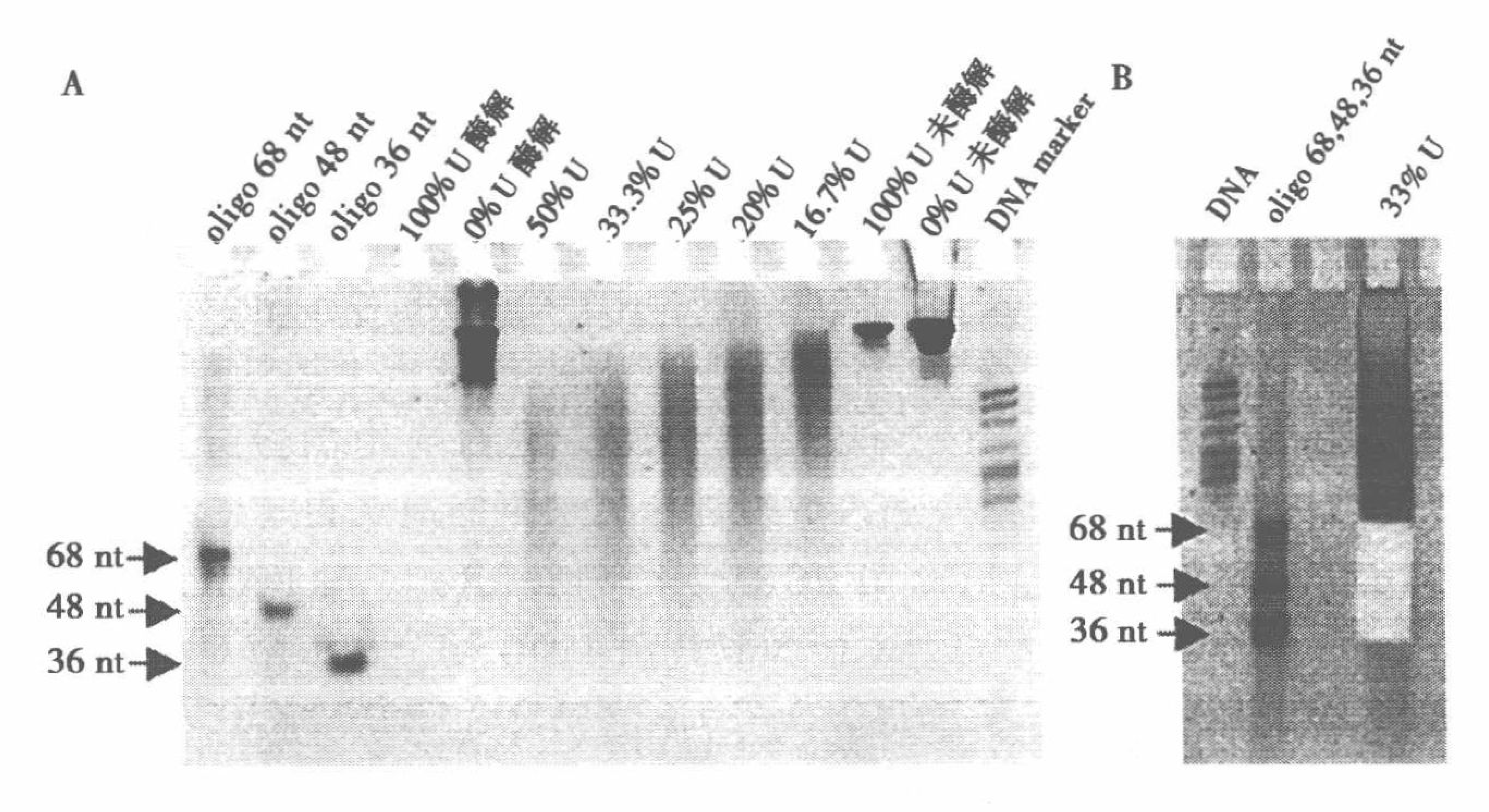


图 10.2　NExT DNA 混编流程的改变尝试。A. 使用 UDG T4-内切核酸酶 V 切割 CAT 野生型基因 PCR 产物的尿素聚丙烯酰胺凝胶电泳结果，使用不同的 dUTP：dTTP 比率对比酶解片段的不同。从左到右依次为条带 1～3，包含 68 个、48 个和 36 个碱基的寡核苷酸；条带 4～10，使用的 dUTP 比率为别为 100%，0，50%，33.3%，25%，20%和 16.7% ，酶解后的产物；条带 11 和 12，使用的 dUTP 比率分别为 100%和 0%，未酶解的产物；条带 13，pBR322/HpaII DNA 分子质量标准。B. 使用 33.3% 比率的 dUTP 获得的 CAT _ Nd10 基因切割片段的制备型聚丙烯酰胺凝胶电泳结果。条带 1，pBR322/HpaII DNA Marker；条带 2，包含 68 个、48 个和 36 个碱基的寡核苷酸混合物；条带 3，切割后酶解前的片段。

10.3.4　尿素变性聚丙烯酰胺凝胶电泳

使用尿素变性聚丙烯酰胺凝胶电泳分析断裂片段的大小分布并纯化特殊大小范围的片段（见 10.3.5）。此步为备选步骤，仅对要求严格控制的断裂反应推荐使用。

(1) 凝胶包含 6.7mol/L 尿素，11%～12%的聚丙烯酰胺和 1×TBE 缓冲液[8]。分析胶的大小为 10cm×8cm×1mm，制备胶的大小为 10cm× 8cm×2mm（见 10.3.5.2）。因尿素降解会影响电泳效果，故要求所有的凝胶都新鲜制备。我们通常使用 31.5g 尿素，29.2ml 包含 30%聚丙烯酰胺，0.8%甲叉双丙烯酰胺（37.5：1）的凝胶储液，17.1ml 水，7.5ml 10×TBE 缓冲液，500μl 10%APS 水溶液和 50μl TEMED；这些足够在一个制胶器中同时制备 9 块分析胶（见注 8）。

(2) 电泳在 56℃进行。我们使用 Hoefer Mighty 小型基础电泳槽（Amersham），配接温度控制水浴锅。

(3) 在 1×TBE 缓冲液中将凝胶预电泳 10min，电压 100V。

(4) 将断裂 DNA 浓缩至 7μl，加至 speed-vac 真空离心挥发哌啶，加 25μl 去离子甲酰胺，在 PCR 仪中 80℃加热 3min（见注 9）。再加入 3μl 60%的蔗糖溶液和 7μl 水帮助上样（见注 10）。最终上样量为 15～20μl（约含 3μg DNA）。

(5) 寡核苷酸混合液和 100bp DNA 标尺可作为分子质量标准（见注 11）。至少一个标尺使用带染料的上样缓冲液。

(6) 上样后 170V 进行电泳，至标尺中的溴酚蓝距边缘还有 0.5cm 时结束（见注 12）。

(7) 在 30ml 1μg/ml 的 EB 溶液（1∶5000 稀释）中染色 5min，紫外灯下观察（见注 13)。

分析型凝胶电泳结果表明，断裂产生片段的大小分布具有固定的峰值，由 dUTP 所占的比率决定。如图 10.1B 所示，多个长度分布带很容易得到，包括很小的和大的片段，分别适用于短的基因和长基因簇的混编。这样的测试只用于最初确定合适的 dUTP 比率以获得目标的片段大小分布。最后确定测试库中 dUTP 使用的最佳比率为 33.3%，且结果的重复性很好。因此，在接下来的定向进化实验中，此步分析电泳可以省略。

10.3.5 片段纯化

经酶解和化学裂解得到的基因片段（见 10.3.3）即可通过硅胶树脂直接从切割反应液中纯化（见 10.3.5.1)，也可通过制备型尿素变性聚丙烯酰胺凝胶电泳纯化（见 10.3.5.2)。

最初我们认为如想得到特定大小的片段必须通过凝胶电泳来纯化，以确保在重组装反应中没有较长片段甚或全长片段的参杂，以免影响碰撞概率。而实际上，如不经多轮的重组装反应，就无法从纯化的片段中直接扩增出全长产物，这表明切割反应进行的十分彻底（见 10.3.7)。因此，只有对片段大小范围有十分严格的要求时才有必要使用凝胶电泳纯化。

10.3.5.1 哌啶切割后直接纯化

(1) 最简便的片段纯化方法就是使用 QiaexⅡ试剂盒（Qiagen）从哌啶切割后的反应液中直接纯化（见 10.3.3)。依据该试剂盒指南进行操作（见注 14)，加入结合缓冲液后中和（约 20μl 3mol/L 乙酸缓冲液)。

(2) 清洗两次后，分两步加入 25μl 洗脱缓冲液洗脱。将两次洗脱液混合。

(3) 离心两次，每次换一个干净的离心管（见注 15)。

10.3.5.2 尿素变性聚丙烯酰胺凝胶电泳纯化

注意此步为可选步骤。尝试了两种方法从凝胶中回收片段。一种是经典的水提取后乙酸/镁离子沉淀法，另一种是使用 Qiagen 公司的 QiaexⅡ试剂盒。但通过染色检测提取完基因后的凝胶，发现两者都存在提取不完全的问题。

(1) 经制备型尿素变性聚丙烯酰胺凝胶电泳后（见 10.3.4，步骤 1)，在低强度紫外灯照射下将目标大小的条带切割下来（图 10.2B)。放入离心管中充分捣碎后加入提取缓冲液（QiaexⅡ提取，步骤 2）或 1ml 水（水提取，步骤 3)，放入热混合器（Eppendorf）中以 37℃，1000r/min 过夜孵育。

(2) 如果使用 QiaexⅡ提取缓冲液提取，见 10.3.5.1 节。

(3) 如使用水提取，加 1/10 体积的乙酸缓冲液，1/100 体积的氯化镁和 1 体积的异丙醇沉淀 DNA；−20℃反应 1h；4℃ 20 000*g* 离心 15min；将沉淀重悬在 50μl 90% 乙醇中，放入热混合器中 37℃，1000r/min 反应 1h；−20℃放置 1h；4℃ 20 000*g* 离心 15min；将沉淀在空气中晾干后，加入 30μl 洗脱缓冲液，37℃，1000r/min 孵育 1h。

为得到胶提取步骤的交叉率，我们设计了3个母本克隆（CAT _ Nd10突变体）的混编检测实验。其中两个克隆在100bp范围内包含1个特异的突变，另一个在100bp范围内包含2个特异突变。混编并经 *Taq* 酶扩增后挑选8个克隆测序，结果表明其中的6个克隆在100bp范围内有一个交叉。这与直接快速纯化得到的交叉率可比（见10.3.5.1和10.3.8）。实验共测序了3851个碱基，发现了12处错误，突变率为0.31%。这样的错误率比快速纯化的错误率高很多，可能是由于UV照射（哪怕是弱的366nm的光源）或是凝胶中的化学修饰引起的。因此我们在最终的NExT DNA混编方法中，去掉了凝胶纯化的步骤，因为额外的工作并未带来益处，反而损失了产量，提高了错误率。

10. 3. 6　纯化片段的定量

为了分析和比较纯化得到的DNA的产量，我们首先使用SYBR GreenⅡ试剂来定量。因为NExT DNA混编方法的重复率很高，我们只定量一次即可（见注16）。通常我们都能获得浓度为40～60ng/μl的DNA片段。

（1）取2μl纯化的DNA片段（见10.3.4），加入50μl 1∶5000稀释的SYBR GreenⅡ溶液染色，此染色液对单链DNA非常敏感（见注17）。

（2）在暗箱中反应5min后检测荧光信号（我们使用96孔的Perkin Elmer LS 50B荧光粉光光度计；见注18）。染料的激发波长是480nm，可在515nm检测发射值。

（3）使用片段大小范围内的标准寡核苷酸，稀释不同的倍数绘制浓度标准曲线，标定片段所含DNA的浓度。

10. 3. 7　基因重组装和扩增

通过内部引物延伸反应从纯化的基因片段重组得到全长基因（见10.3.5），其间不断提高退火温度并最好使用具有校正功能的DNA聚合酶，如Vent（图10.3）。在内部引物延伸PCR过程中，片段之间可相互作为引物，因此每个循环反应后都会加长，直到最终得到全长基因。最后，通过常规PCR反应，加入两个末端引物对组装反应的产物进行扩增，构建克隆送测序。在方法建立过程中，通过琼脂糖凝胶电泳不断检测重组装反应结果（图10.3）。重组装反应在进行若干轮后停止，所得产物在此基础上进行PCR扩增。最后一步反应的基本原理与其他的基因组装进程相同[3]，但有很重要的改变。尽管经历了剧烈的化学切割条件，使用了高保真DNA聚合酶的组装反应仍很高效。

（1）取出约2μg的纯化DNA片段（见10.3.5）进行重组装。事实上如使用Vent聚合酶，少于1μg的DNA片段都够用。通常我们都使用20～25μl的片段溶液，无须检测浓度。

（2）在DNA片段中加入4μl（见注19）10mmol/L的dATP、dTTP、dCTP和dGTP的混合物（终浓度800μmol/L），以及4 U Vent DNA聚合酶（NEB），1～4μl 25mmol/L的硫酸镁，5μl的10×反应缓冲液，加水定溶至50μl。

（3）循环反应为94℃ 3min；92℃变性30s，30℃退火60s，每轮增加1℃（冷却率1℃/s），72℃延伸1min，每轮多加4s，共进36个循环；最终72℃再延伸3min。

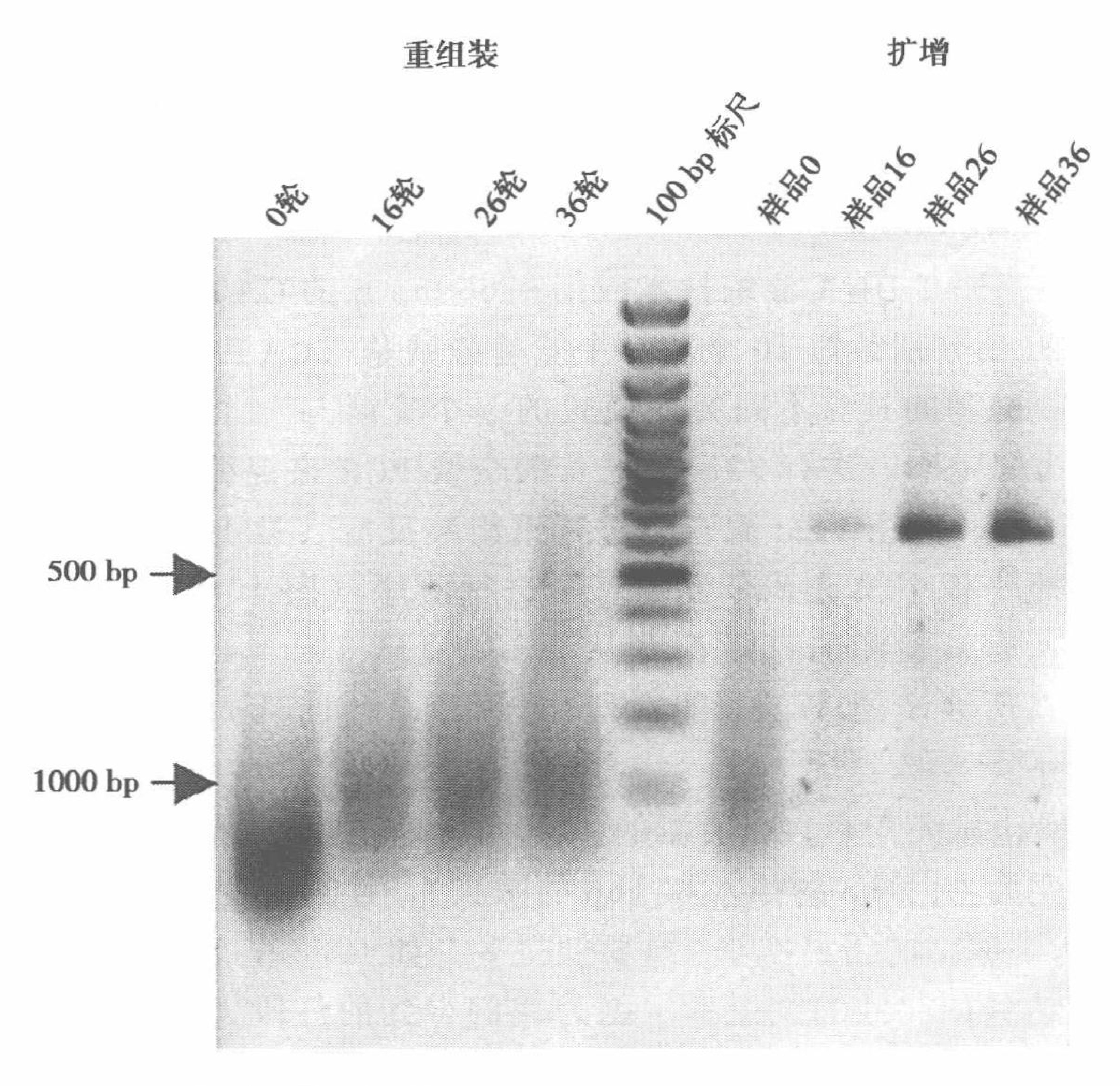


图 10.3　混编基因的重组装和扩增。包含 33.3%U 的混编的 CAT-Nd10_Cd9DNA 基因片段的 1%琼脂糖凝胶电泳结果，使用 vent DNA 聚合酶进行重组装后再通过 Taq 聚合酶扩增。条带 1：未经重组装 PCR 的片段；条带 2：经 16 个循环重组装的片段；条带 3：经 26 个循环重组装的片段；条带 4：经 36 个循环重组装的片段；条带 5：100bp DNA 标尺；条带 6 未经重组装的片段直接进行 PCR 扩增；条带 7：经 16 轮重组装的片段进行 PCR 扩增；条带 8：经 26 轮重组装的片段进行 PCR 扩增；条带 9：经 36 轮重组装的片段进行 PCR 扩增。

(4) 取重组装产物 10μl（见注 20）使用基因的合适引物（25pmol 引物，0.2mmo/L dNTP，25 个循环，40s 的延伸时间），进行常规 PCR 反应。

(5) 使用适当的限制性酶切位点将扩增的基因克隆到载体上。

据我们所知，大多数已发表的片段重组装过程都使用 *Taq* 酶。Vent DNA 聚合酶的 3′→5′校对核酸外切酶活性是基于其仅在 3′端移除合成链上错配的核苷酸，直到聚合反应可以从退火的末端开始。这意味着核苷酸链上的点突变，即便是紧邻 3′端的位置，都不会有问题[19]。为了比较这两种聚合酶的作用，分别使用 *Taq* 和 Vent 酶平行进行两组重组装实验，均使用相同的片段库作为起始。适量的重组装产物作为扩增 PCR 的模板，当使用 *Taq* 酶时，产物经琼脂糖凝胶电泳只有一条带，然后通过图像分析定量。结果，使用 Vent 酶的得到 DNA 产量为 *Taq* 酶的 35 倍。有意思的是，将使用 *Taq* 酶得到的 6988 个碱基测序，发现只有 5 个额外的突变（0.075%）。因此，在现有序列的统计误差内，使用 *Taq* 和 Vent 酶进行 NExT DNA 混编产生的错误率基本是相同的（见 10.3.8）。关于两者产量的显著差别，可用校正酶的性质来解释。以 Vent 酶为例，该酶具有链置换活性[20]，这可能会帮助许多杂交反应的进行，另外 Vent 酶在 95℃的半衰期约为 8h，相对于 *Taq* 酶的 1.6h[20]，更适合长时间的重组装反应。另外，*Taq* 酶习惯在 3′端多加额外的 dATP[21]，这也是妨碍重组装反应进行的可能原因。除

了上述因素，还发现 *Taq* 酶虽然适合片段的重组装反应，但其重组装的产物作为下一步扩增反应的模板，很难使用具有校正功能的 DNA 聚合酶进行扩增。

10.3.8　交叉率、片段的平均长度及突变率分析

将上面描述的 NExT DNA 混编技术应用在 600bp 长的 CAT 基因的定向进化研究中，其中 N 端和 C 端分别截短 10 个和 9 个氨基酸残基（CAT _ Nd10 _ Cd9）。实验中，挑选 12～383 碱基间具有不同突变模式的 5 个克隆与一个用于回交的截短野生型克隆相混合作为固定库，进行 33.3％尿苷酸交换 PCR 来混编。从溶液中直接纯化片段（见 10.3.5.1），再经 Vent 聚合酶进行重组装反应（见 10.3.7 节）。在没有抗性选择压力的对照平板中挑选 8 个混编克隆进行测序（图 10.4A）。这些克隆独特的突变模式显示所有被检测的克隆都是从 2（如克隆 1）～4 个（如克隆 4）母本克隆演化来的。在所涉及的 372bp 的片段中，每 93～186bp 就有一个交叉，平均片段长度为 114bp。

通过测序结果计算突变率。在所测的 4425 个碱基中，发现了 4 个突变（一个 A 变为 G，一个 T 变为 C，1bp 的插入和 1bp 的缺失），计算突变率为 0.09％。这比之前报道的 DNA 酶混编[2]的错误率 0.7％要低得多。如 10.3.7 节中所述，并不是只有 Vent 聚合酶才能达到如此低的突变率。因为我们得到的片段大小分布和交叉率均与 DNA 酶混编的实验结果相当，因此分析前面报道的高错误率可能是由 DNA 酶解及凝胶观测时的紫外损伤引起的，而与片段大小和聚合酶种类无关。低突变率对长片段 DNA 的混编尤为重要，因为我们要尽可能地避免将功能异常的或不需要的分子引入基因库中。

进一步的实验中，挑选 4 个截短 CAT 基因（CAT _ Cd26）母本进行了上述相同的混编实验，它们均包含一个突变，分布在 9～575 碱基，最终挑选 5 个克隆测序（图 10.4B）。可检测的片段平均长度为 86bp，其中包括许多短的片段。在“对照 5”克隆中，突变位置分别相隔仅 6 个碱基（324 和 330 位置），12 个碱基（364 和 376 位置）及 11 个碱基（404 和 415 位置）的突变体都被分开。这次实验得到片段的平均长度比之前的实验要小，这是因为更多的突变利于片段长度的检测。至于许多短片段的获得可能原因有两点：使用的 QiaexⅡ试剂盒可高效的回收短片段 DNA，或者基因扩增时的交叉是一个复杂的过程，包含如 PCR 基础上的链交换。

10.3.9　NExT 片段断裂一预测程序

因为 dUTP 的掺入及产生的片段断裂都基于可推导的原则，我们开发了名为 NExTProg 的计算程序[22]。该程序可预测双链 DNA 的 NExT 断裂模式，使得研究人员不需要实验就可设计合适的 dUTP：dTTP 比率。该程序通过读入 DNA 序列文件和 dUTP：dTTP 比值，就可算出所有可能的片段、它们出现的可能性及相对分布。给定 DNA 的互补链由程序自动生成并加入计算。计算后的结果以柱型统计图表形式显示，所有数据都可以列表形式输出以备使用（图 10.5A）。当需要设定片段大小的上下限时（如凝胶纯化所需），程序可计算材料损失的可能性并调整每个片段的相对分布值。

A

trunc. gene position	12	52	53	55	73	74	87	116	127	194	231	301	326	383	母本克隆最小值	每100bp的突变率
original nt	C	G	T	G	C	A	G	A	C	T	T	G	T	T		
clones used (CAT Nd10 Cd9)																
parent wt																
parent a		A	C			T	A	G			G	C	C	C		
parent b	A	A	C	A		T			T			C	C	C		
parent c		A	C		T	T		G			G	C	C	C		
parent d	A	A	C	A		T			T	C		C	C	C		
parent e	A	A	C	A		T						C	C	C		
clones obtained																
control 1												C	C	C	2	0.00
control 2									T						3	0.18
control 3								G						C	4	0.18
control 4	A										G	C	C		4	0.00
control 5		A	C	A		T	A		T			C	C	C	4	0.00
control 6													C		3	0.18
control 7		A	C		T	T						C	C		3	0.00
control 8		A	C	A		T									3	0.20

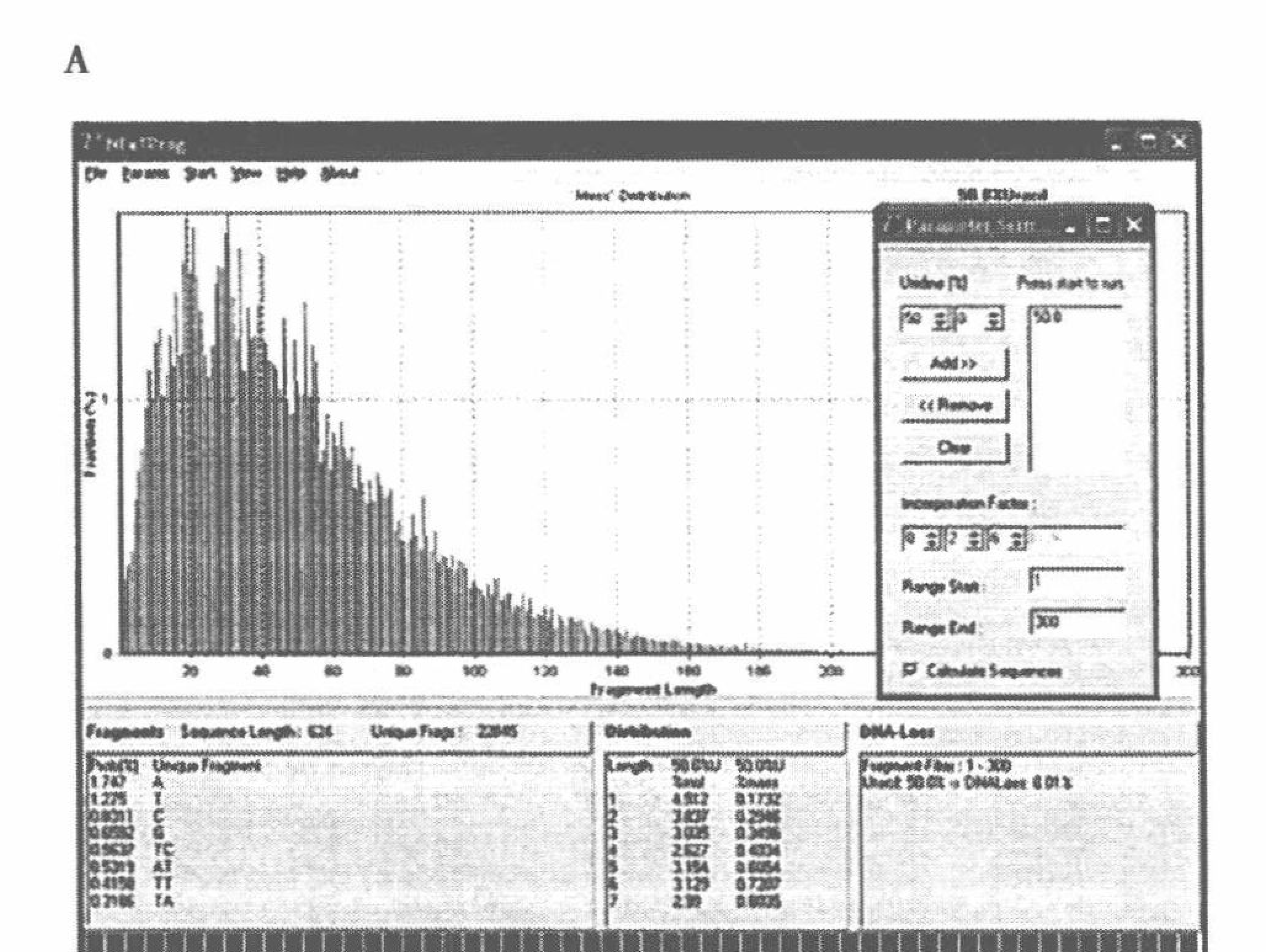
B

gene position	9	11	25	36	55	73	104	113	132	150	163	176	180	181	183	193	227	255	264	281	300
original nt	A	A	A	A	T	T	A	A	T	A	T	T	T	C	T	C	C	G	T	T	A
clones used (CAT Cd26)																					
parent a					A		G		C	T			A					T			T
parent b						C					C								C	A	
parent c	G	T	T	T				G				C		T							
parent d																					
clones obtained																					
control 1							G		C	T			A				T	T			
control 2							G		C	T			A					T			T
control 3								G	C	T			A								T
control 4				T			G		C	T			A								T
control 5					A		G			T			A								

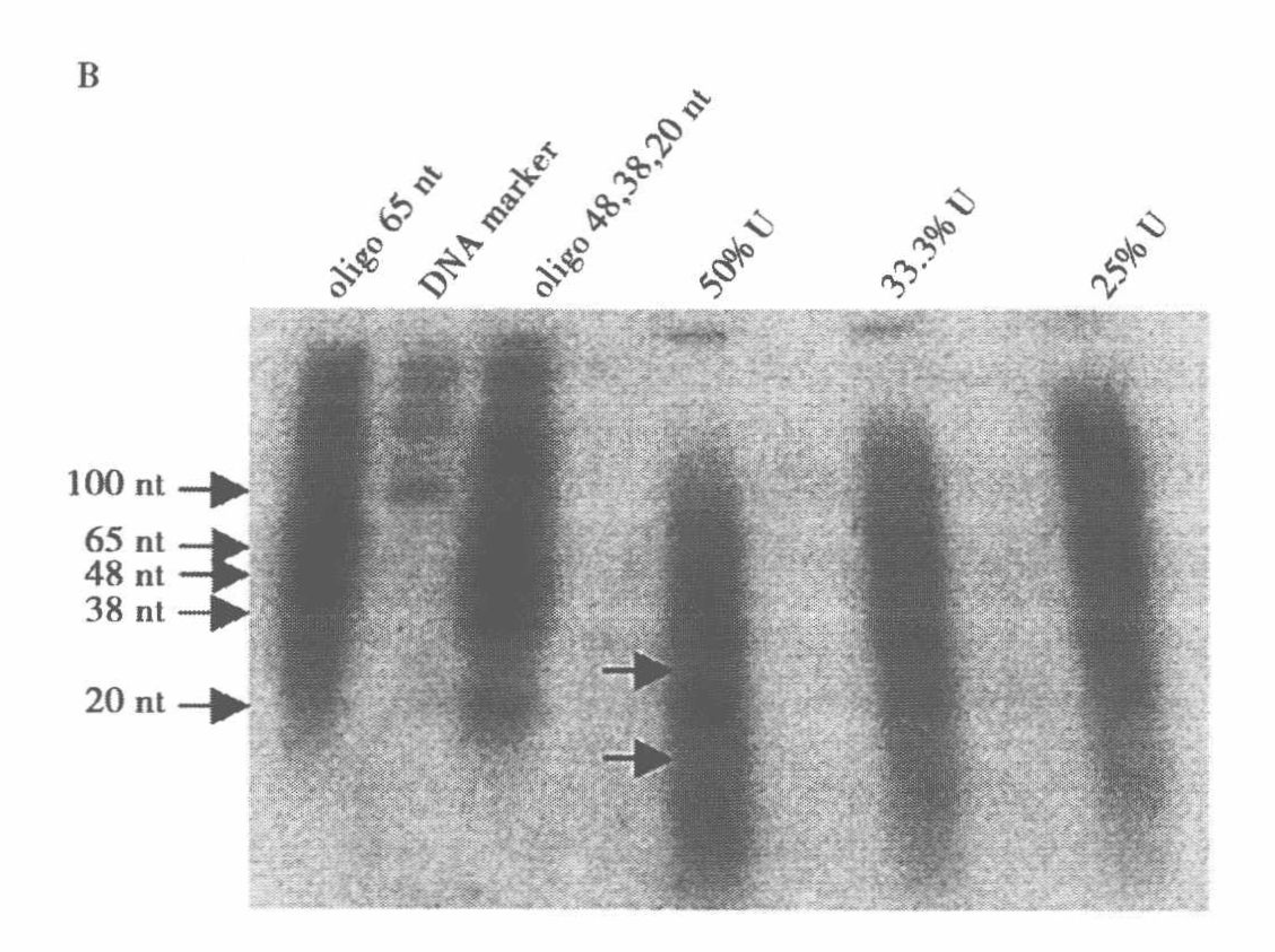
gene position	302	324	330	336	364	366	376	383	404	415	422	428	430	436	456	458	492	521	544	549	556	575	母本克隆最小值	每100bp的突变率
original nt	C	A	C	T	C	A	T	G	C	A	A	T	T	T	T	T	C	G	A	G	C	A		
clones used (CAT Cd26)																								
parent a									T	T								T						
parent b							C				G			C										
parent c	T			C	T		G					C	A		G	C	T				T			
parent d																			G					
clones obtained																								
control 1									T	T													7	0.16
control 2									T	T													3	0.00
control 3									T	T													6	0.16
control 4																							7	0.33
control 5					T				T														10	0.00

图 10.4 交换频率和片段长度分析。A. CAT_Nd10_Cd9 基因突变库进行 NExT DNA 混编后的测序结果，其中使用直接从溶液中纯化的片段，并用具有校对功能的 DNA 聚合酶进行重组装。每个克隆都大概测了 500～571 个碱基。混编的检测实验中，包含 CAT 野生型片段 26ng(52%)及每种突变体片段4.8ng(9.6%)进行含 33.3%尿苷酸的交换 PCR，下面的表格列出了未经选择压力得到的克隆的序列，主要包括混编的突变位置、由突变模式推测出的最小母本克隆数以及上表中未列出的额外突变的引入率。平均下来，372bp 的片段共包含 3.25 个母本克隆。由于加入了过量的野生型基因，实际包含母本克隆的数目应比列出的最小值要高。B. 等量混合 4 个 CAT_Cd26 母本克隆进行 NExT DNA 混编后的测序结果。每个母本克隆都用不同的阴影表示，可作为混编克隆的片段识别。没有阴影的部分表示可能是从多个母本获得的片段。检测到片段的平均长度为 86bp。

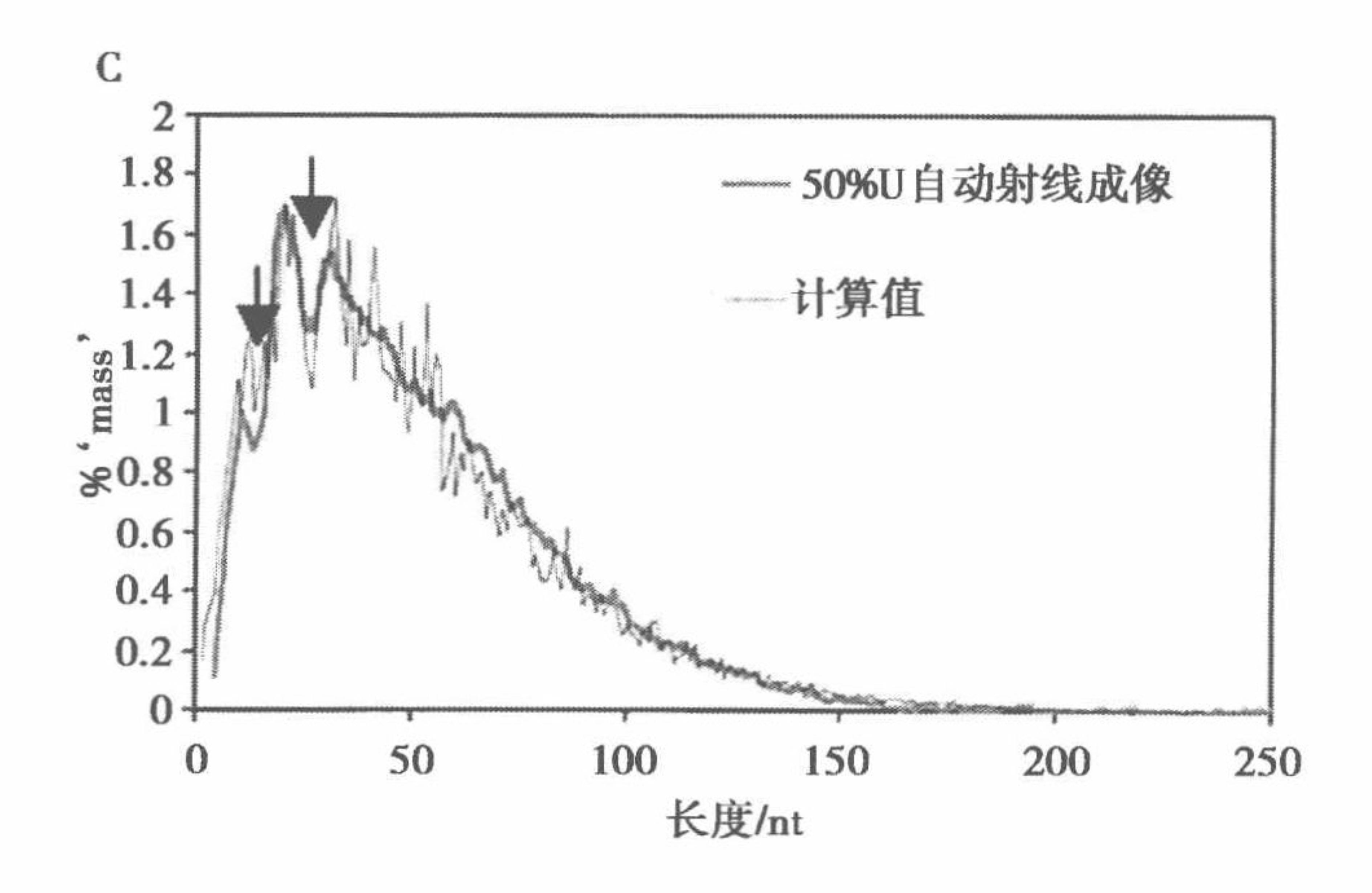
A



B


oligo 65 nt
DNA marker
oligo 48,38,20 nt
50% U
33.3% U
25% U
100 nt
65 nt
48 nt
38 nt
20 nt

C


50%U自动射线成像
计算值
% 'mass'
长度/nt
0
50
100
150
200
250
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2

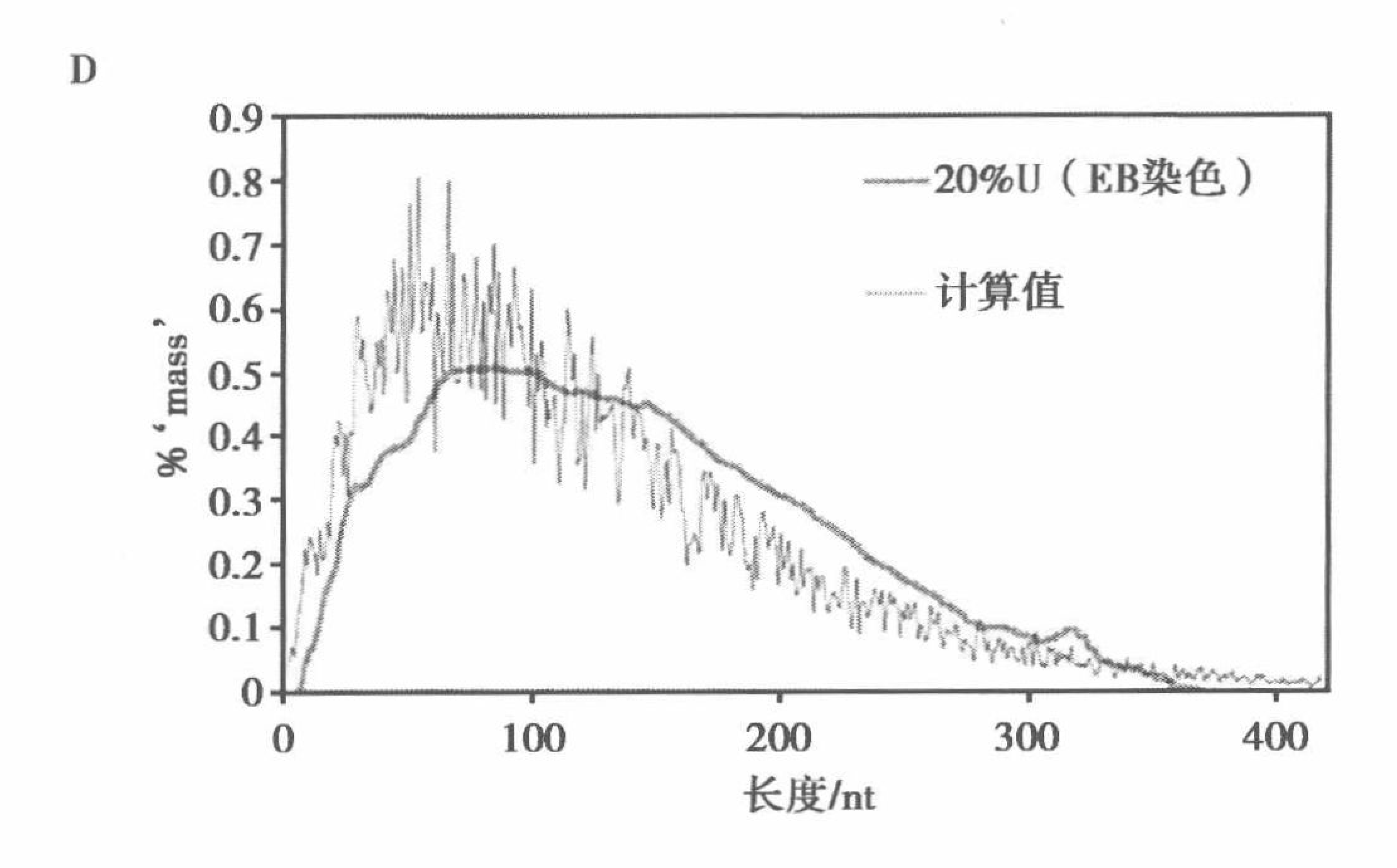


图 10.5 NExT 断裂实验中计算结果与放射性标记及 EB 染色的检测结果的对照。A. NexTProg 1.0 的图形窗口[22]。该程序通过读入 DNA 序列计算所有可能产生的片段。程序分别给出片段长度对均一化的分子比率或 mass 值的列表，定义为核苷酸的数目。所有片段的序列均可生成，而相同的序列被合并在一起，用于后面的组合计算。B. 放射性标记 DNA 样品的尿素变性聚丙烯酰胺凝胶电泳结果。条带 1～3：标记^{32}P-ATP 的寡核苷酸分子质量标准；条带 4～6：CAT _ Cd26 基因（624bp）片段，交换 PCR 中分别使用标示量的尿苷酸及^{32}P-CTP（每条带都包含 0.3 μCi)。通过荧光描仪（Kodak）进行放射性自显影后再通过荧光成像仪（BioRad Fx）读数。注意各条带下部 1/3 段的差别，分别对应断裂产生的序列特异峰。C. 测量和计算得到的片段分布，用于确定交换 PCR 时尿苷对胸腺嘧啶的掺入率。灰色的线代表 B 图中的 50%U 条带放射性的线密度曲线，根据分子质量标准由相对迁移距离转换为核苷酸长度，再通过平均相应的值设为整数进行均一化。黑色线代表 NexTProg 程序计算的同一基因的片段“mass”值分布，设定 50%U 和 0.26 的掺入率，引入的均方根偏差最小。D. 灰色线代表 20%U（图 1B，C）反应的线密度曲线，经 EB 染色后转换为片段长度并均一化。黑色线为 NexTProg 程序的计算值，设定 20%U 和 0.26 的掺入率。注意短的单链寡核苷酸使用 EB 染色效果不好，因此均一化的曲线中长片段的显示量比实际量要高。

10.3.9.1 数学原理

（1）设定给定 DNA 序列中的一个胸腺嘧啶被尿嘧啶替代的概率为 p。

（2）当两个胸腺嘧啶都被尿嘧啶替代时，就生成两者之间的片段，其概率为 $p \times p$。

（3）只有当上述两个胸腺嘧啶间的 n 个胸腺嘧啶都不被替代时，片段才能存在，得到的概率为 $p \times p \times (1 - p)^n$。

（4）对于包含 1 个或 2 个 DNA 末端的片段，$p \times p$ 概率中的 1 个或 2 个 p 设为 1（见注 21）。

（5）为比较断裂结果，我们定义片段概率为除以所有值的和。

我们的计算方法与先前的计算都不同[23]，因为我们考虑到片段的两个末端都需要产生，但不考虑诸如序列偏倚性和不确定的酶解条件等问题，这些会严重阻碍对 DNA 酶酶解的计算。

对于包含 x 个尿苷酸的基因，最多可产生 $\sum_{i=1}^{x+1} i$ 个可能片段［根据高斯定理，当 x 为奇数时，等于 $(x+1) \times (x+2)/2$；见注 22］。因此，对于一个典型的包含 250 个

胸腺嘧啶和 240 个腺嘌呤（在互补链中也计算为胸腺嘧啶）1000bp 长的基因，很快程序计算产生31 626+29 161=60 787 个片段，包括它们的概率和序列。

因为大多数用户可能都对结果概貌感兴趣，该程序将所有具有相同大小的片段汇集在一起，将它们的概率加和，给出占所有概率加和的百分比对长度的分布，在程序中以“%mol”表示。为了将电泳条带可视化，通过某一片段的概率乘以其长度计算“mass”分布。这些值被均一化，代表碱基比率，定义为“mass”，如在碱基对中固定长度，程序中显示为“%mass”。片段序列的输出文件按出现可能性由大到小列出了所有的片段。同样的序列只列出一次，概率为它们的加和。

10.3.9.2　程序的校准及与实验结果的对比

在比较测量和计算的结果之前，有一个非常重要的值要考虑：就是交换 PCR 时聚合酶掺入尿苷酸对胸苷酸的比率。这个值可能不仅取决于 dUTP∶dTTP 的值，也取决于使用的聚合酶类型、核苷酸的绝对浓度以及使用的缓冲液。因此当使用程序时，需要设定该值。尿苷酸的掺入量可用下列方法测量：直接的方法是通过放射性标记的 dUTP，计算 PCR 产物的放射能；间接的方法是定量分析断裂图谱。我们选择后者，因为它同时提供了与我们的软件相似的基本原则。但需要注意这种对资料的双重利用只有在断裂完全的情况下才有效。

为定量分析裂解实验，我们使用常用的条件（标准 *Taq* 聚合酶[24]，200μmol/L dNTPs）。

（1）在尿苷酸交换 PCR 反应中（见 10.3.2，见注 2），使用^{32}P-dCTP 标记的 DNA。

（2）进行变性聚丙烯酰胺凝胶电泳（见 10.3.4）。

（3）通过成像仪对凝胶进行放射能成像（见注 23）。这避免了因小片段的低效染色造成的信号变形。

（4）将每条带都通过图像分析软件，如 QuantityOne（BioRad）或 NIH Image/Scion Image/imagej）生成线密度曲线。

（5）从线密度曲线中计算出片段和放射标记的分子质量标准的相对迁移值。

（6）将代表标记物相对距离和长度的可变参数 a 和 b 代入方程：相对距离＝a×ln（碱基对长度）＋b。

（7）用上述公式将片段的相对距离值换算为核苷酸长度。

（8）通过结合整数的核苷酸长度值和将相应的强度信号平均化，将强度信号/连续长度分布转换为强度信号/整数（只列出分散的寡核苷酸长度）。

（9）通过将每个单独的强度除以强度总和，将分布均一化。

上述变性胶的放射能成像图见图 10.5B。强度均一化后表示为“%mass”（见图 10.5C）。将这一实验与程序计算结果相比较确定的尿苷酸对胸苷酸的相对掺入值。程序计算中，设置的掺入值为 0.2～0.3，按 0.01 增长。当设置值为 0.26 时，实验和计算结果最为吻合，其中根据测量和均方根分析预测的片段大小范围为 10～150 个碱基，而计算得到的可靠片段长度为 4～200 个碱基（图 10.5C）。当使用分子质量标准的范围时（20～100 个碱基），因子最佳的设置值为 0.25。很显然这些图谱重叠得很好，并且电泳条带中的峰和点都可以很好地解释（图 10.5B、图 10.5C）。重要的是，*y* 轴的值

落在了不需进一步调整的位置。我们确信尿苷酸的掺入率是确定的，并且没有不完全断裂造成的影响，因为实验和计算得到的结果几乎一模一样，并且尝试了许多实验条件检测断裂情况。使用 33.3%U 断裂的放射活性测量值与计算值也吻合得很好。但使用 25%U 断裂的放射活性检测结果中产生了相当量的大于 100bp 标准的片段，并且出现了一些因标尺问题引起的偏离。然而同样使用掺入率 0.26，与使用 EB 染色后的断裂结果相比较，考虑到凝胶染色的长度依赖性，结果十分吻合（图 10.5D），其中 EB 染色因有合适的分子质量标准，对于较长片段更为准确。因为针对放射活性检测和 EB 染色的尿苷酸掺入 PCR 分别使用了不同的缓冲液（见 10.2.2），多种 PCR 增强剂，如硫酸铵、吐温-20 或 Triton X-100 的加入对 dUTP 对 dTTP 的掺入率并没有显著的影响。

10.4 注

1. 尽管 NexT DNA 混编方法的整体错误率已经很低，尿苷酸交换 PCR 中仍然推荐使用高保真聚合酶。但是，须谨慎选择该聚合酶的种类以使其能够有效地掺入尿苷而不是在模板的尿苷酸位点停滞。Vent exo^{+}（NEB）以及非尿苷酸停滞突变体的 *pfu* 聚合酶（Stratagene）都被报道可以掺入尿苷。

2. 很小的碎片能够被放射性同位素标记显示。这种情况下，加入 0.5μl 的 3.3μmol/L ^{32}P-dCTP 溶液（约 0.5μCi）到尿苷交换 PCR 混合物。对于放射性实验，使用 PCR 清洁试剂盒之前不用经过琼脂凝胶电泳步骤（见 10.3.2.4），因为非放射性模板不影响放射性片段的显示。

3. 由于热循环仪最适工作体积的限定，我们用了 4 个 50μl 试管的尿苷交换 PCR 反应。

4. 需要分离含有尿嘧啶的产物及不含尿嘧啶的模板，以确保没有模板被保留到下一步。

5. 我们更倾向于利用两个柱层析进行平行实验以节省时间。这样，我们在两个试管中分别进行尿苷交换 PCR（见 10.3.3）的酶及化学切割实验以保证我们的热循环仪在合适的工作体积内。样品可以在 10.3.5.1 第二步或者 10.3.4 第四步组合起来进行片段纯化。

6. 酶解 2h 或者更长时间产生同样结果，表示酶解过程的选择性及一致性。

7. 哌啶有毒，需要在通风厨中操作。所有其他含有哌啶的溶液，如 Qiaex Ⅱ试剂盒中（见 10.3.5.1）用过的俘获缓冲液，需要谨慎处理，弃于密闭试管中。

8. 我们制备的凝胶最多用 3d。加入的尿素粉末影响体积，需要被考虑到终浓度中。体积为 9 块凝胶的溶液被用于制备 8 块凝胶。

9. 哌啶影响片段的电泳性质，因此，应该尽最大可能使其挥发。

10. 对于放射性实验，仅仅使用 7μl DNA-甲酰胺样品，补充加入 3μl 60%的蔗糖溶液和 7μl H_2O；混匀后取 15μl 样品加入凝胶中。

11. 对于放射性实验，寡核苷酸和 DNA 分子质量梯度被^{32}P-γATP 磷酸化，然后用分子筛纯化去除多余的^{32}P-γATP。我们利用 Sephadex G-50 柱料，但是任何其他核酸去除试剂盒都适用。

12. 为避免染料与 DNA 片段叠合，我们只是在分子标记中用了溴酚蓝而没有在样品中用。

13. 染色孵育时间不要超过 5min，否则较小的片段会开始从凝胶中跑出。因此，用 SYBR green Ⅱ 染色比较困难，因为这种染剂染色需要孵育 5～10min。请注意尿素凝胶在紫外光下很快变白。

14. 我们发现，对于避免损失小片段，Qiaex Ⅱ 试剂盒是最适用的。尽管厂家推荐使用的 DNA 片段范围是大于 40 个碱基，我们发现此试剂盒可以很好地回收大约 20 个碱基的片段。

15. 由于试剂盒基质会结合 DNA 从而影响重组装反应，因此离心两次来确保 DNA 片段不会被微量的基质材料污染是很必要的。

16. 一般来说，如果我们开始就用 7μg 或更多的尿苷交换 PCR 反应产物进行实验，就不必检测 DNA 片段的浓度。

17. 供应商推荐使用 1∶10 000 的稀释倍数，我们发现对于小的和低浓度的 DNA 片段 1∶5000 的稀释效果更好。

18. 最好使用白色的 96 孔 SBS 板，因为白色板会增强荧光。

19. 我们选用了较高量的 dNTP，因为在长达 4h 的重组装反应中 dNTP 可能降解。

20. 选择此体积以确保多样性。

21. 基因中尿苷掺入的所有可能类型的总和是 1，但是所有片段的概率总和要大一些，因为每种类型可以产生好几种片段。

22. 这个计算包括长度为零碱基的片段，当两个尿嘧啶紧挨在一起时会产生这种情况。

23. 同位素放射自显影的胶可应用 X 射线光片显影。但是，现代化影像设备可以给出更大的线性测量范围。

致谢　我们感谢 Susanne Knall 女士在实验中给予的帮助，Gregor Zipf 编写 NexTProg 程序，与 Hubert Bernanser 的富有成效的讨论，以及 Jody Mason 严谨评判性地阅读原稿。K. M. A. 由 Deutsche Forschungsgemainschaft，Emmy Noether-Programm（Ar373）等基金支持。

参考文献

1. Neylon, C. (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolu-tion. *Nucleic Acids Res* 32, 1448-1459.
2. Stemmer, W. P (1994) DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* 91, 10747-10751.
3. Stemmer, W. P. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389-391.
4. Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* 16, 258-261.
5. Shao, Z., Zhao, H., Giver, L., and Arnold, F. H. (1998) Random-priming in vitro recombination: an effective tool for directed evolution. *Nucleic Acids Res.* 26, 681-683.
6. Cadwell, R. C. and Joyce, G. F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* 2, 28-33.
7. Arndt, K. M., Müller, K. M., and Plückthun, A. (2001) Helix-stabilized Fv (hsFv) antibody fragments: substituting the constant domains of a Fab fragment for a heterodimeric coiled-coil domain. *J. Mol. Biol.* 312, 221-228.

8. Sambrook, J. and Russel, D. W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

9. Patel, P. H. and Loeb, L. A. (2000) Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.* 275, 40266-40272.

10. Boiteux, S., O'Connor, T. R., and Laval, J. (1987) Formamidopyrimidine-DNA glycosylase of *Escherichia coli*: cloning and sequencing of the fpg structural gene and overproduction of the protein. *Embo J.* 6, 3177-3183.

11. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) *DNA Repair and Mutagenesis*. American Society of Microbiology, Washington, DC.

12. Drohat, A. C., Jagadeesh, J., Ferguson, E., and Stivers, J. T. (1999) Role of electrophilic and general base catalysis in the mechanism of *Escherichia coli* uracil DNA glycosylase. *Biochemistry* 38, 11866-11875.

13. Maxam, A. M. and Gilbert, W. (1980) Sequencing end-labeled DNA with basespecific chemical cleavages. *Methods Enzymol.* 65, 499-560.

14. Ljungquist, S. (1977) A new endonuclease from *Escherichia coli* acting at apurinic sites in DNA. *J, Biol. Chem.* 252, 2808-2814.

15. Richardson, C. C. and Kornberg, A. (1964) A deoxyribonucleic acid phosphataseexonuclease from *Escherichia coli*. I. Purification of the enzyme and characterization of the phosphatase activity. *J. Biol. Chem.* 239, 242-250.

16. Dodson, M. L., Schrock, R. D., 3rd, and Lloyd, R. S. (1993) Evidence for an imino intermediate in the T4 endonuclease V reaction. *Biochemistry* 32, 8284-8290.

17. Levin, J. D. and Demple, B. (1990) Analysis of class II (hydrolytic) and class I (beta-lyase) apurinic/apyrimidinic endonucleases with a synthetic DNA substrate. *Nucleic Acids Res.* 18, 5069-5075.

18. Miyazaki, K. (2002) Random DNA fragmentation with endonuclease V: application to DNA shuffling. *Nucleic Acids Res.* 30, e 139.

19. Mattila, P., Korpela, J., Tenkanen, T., and Pitkanen, K. (1991) Fidelity of DNA synthesis by the Thermococcus litoralis DNA polymerase—an extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res.* 19, 4967-4973.

20. Kong, H., Kucera, R. B., and Jack, W. E. (1993) Characterization of a DNA polymerase from the hyperthermophile archaea Thermococcus litoralis. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *J. Biol. Chem.* 268, 1965-1975.

21. Clark, J. M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* 16, 9677-9686.

22. Müller, K. M. and Zipf, G. (2004) NExTProg 1.0, download available at http://www.molbiotech.uni-freiburg.de/next or http://www.ATG-biosynthetics.com.

23. Moore, G. L. and Maranas, C. D. (2000) Modeling DNA mutation and recombination for directed evolution experiments. *J. Theor. Biol.* 205, 483-503.

24. Kaledin, A. S., Sliusarenko, A. G., and Gorodetskii, S. I. (1980) [Isolation and properties of DNA polymerase from extreme thermophylic bacteria Thermus aquaticus YT-1]. *Biokhimiia* 45, 644-651.

11 简并寡核苷酸基因混编

PeterL. Bergquist and Moreland D. Gibbs

概要 蛋白酶生化性质的改善可以通过对该酶基因的错掺突变和DNA混编方法实现。混编技术可用于同一基因的一组突变体，或者对相关家族基因的片段进行新的组合，产生嵌合突变基因产物。但该方法有一个缺点，就是在混编重组过程中，突变库中会存留大量并未发生重组的（“亲本”）片段。现在，我们可以通过设计简并引物，利用简并寡核苷酸基因改组方法来提高发生重组基因的产量，减少未重组“亲本”基因的产量。在实验过程中，该方法既可避免混编前使用核酸内切酶将基因切成片段，又可对基因特定片段进行随机突变。这里我们详细描述了这种方法如何应用于不同GC含量的β-木聚糖酶家族基因。

关键词 聚合酶链反应；引物延伸；简并引物；体外进化；基因混编；互补简并端引物

11.1 引　言

直到近来，创建组合文库的最流行的方法是通过引入点突变产生一系列进化序列的递归策略。对于体外进化来说，比起递归突变的方法，重组聚合酶链反应（PCR；基因混编）有着实践和理论上的优势[1~3]。该方法通过重组融合多重点突变的序列和野生型的序列，可快速得到合适的突变蛋白。相关家族基因重组通过将这些基因的片段进行混编，重组后利用PCR技术，得到新的基因。该方法的PCR组装步骤主要是利用同源重组。因为“交叉点”经常发生在序列很类似的区域，所以这些方法都需要基因之间有相对高的序列同源性。

如果进行重组的序列之间相似性比较低，则主要产生未重组的“亲本”基因，需要比较费力地寻找发生重组的基因[4,5]。Kichuchi等描述过一种利用“亲本”基因上的特异限制性酶切位点来进行基因重组的方法[5]。该方法可以使基因发生混编的频率增高。

我们分离到一种嗜热的β-木聚糖酶，它在纸浆漂白方面有着优越的表现[6]。我们希望通过基因突变的方法获得一种更稳定的以及变化最适pH情况下的β-木聚糖酶。我们首先运用易错聚合酶链反应（error-prone PCR）和错掺突变，然后进行基因混编，这种方法只能在有限的序列上产生突变，并且需要花费很大精力去筛选这些突变体。另外，通过DNA酶I处理的相关家族基因片段，然后进行混编（家族混编），发现主要产物是野生型的序列。为了避免上述情况，通过方法上的改进，我们发明了一种技术，可以使基因重组发生在序列相似性不高和GC含量差别很大的片段上，且可以显著减少未发生重组的基因产量。更进一步，可以控制引

物延伸的条件，从而使子代基因与亲代基因不同，我们将这种方法称为简并寡核苷酸基因混编（degenerate oligonucleotide gene shuffling，DOGS；参考文献［7］），它也可与其他突变技术一起使用。

11.2 材　料

（1）基因源：家族 11 木聚糖酶基因来自以下菌株：*Dictyoglomus thermophilum* strain Rt46B.1 *xynB*[8]；*Clostridium stercorarium xynB*[9]；*Bacillus* sp. strain V1～4[10]；*Caldicellulosiruptor* sp. strain Rt69B.1 *xynD*[11]；*Clostridium thermocellum xynV*[12]；*Streptomyces roseiscleroticus xyl3*[13]。每个基因都是从相应基因组中扩增出来的。

（2）*Pfx* 聚合酶（Invitrogen，Victoria，澳大利亚）。

（3）pBSII KS 质粒（Stratagene，San Diego，美国加州）。

（4）碱性磷酸酶（Roche Diagnostics Australia，NSW，澳大利亚）。

（5）大肠杆菌 DH5α。

（6）通用缓冲液：50mmol/L 磷酸、50mmol/L 硼酸、50mmol/L 的乙酸，用氢氧化钠调 pH。

（7）覆盖液：0.5%桦木木聚糖（birchwood xylan）(Sigma-Aldrich，Sydney，澳大利亚)，0.5%琼脂糖溶解在 120mmol/L 的通用缓冲液中，pH6.5。混合，然后高压灭菌，最后冷却到大约 50℃。

（8）刚果红溶液（Congo Red）：1% Congo Red（Sigma-Aladrich，Sydney，澳大利亚）溶于水。加入氢氧化钠到大约 5mmol/L，保证溶液的 pH 呈微碱性。

（9）脱色液：1mol/L 氯化钠、5mmol/L 氢氧化钠。

（10）蛋白质抽提试剂 BPER II (Pierce Chemical Company，Rockford，美国伊力诺伊州)。

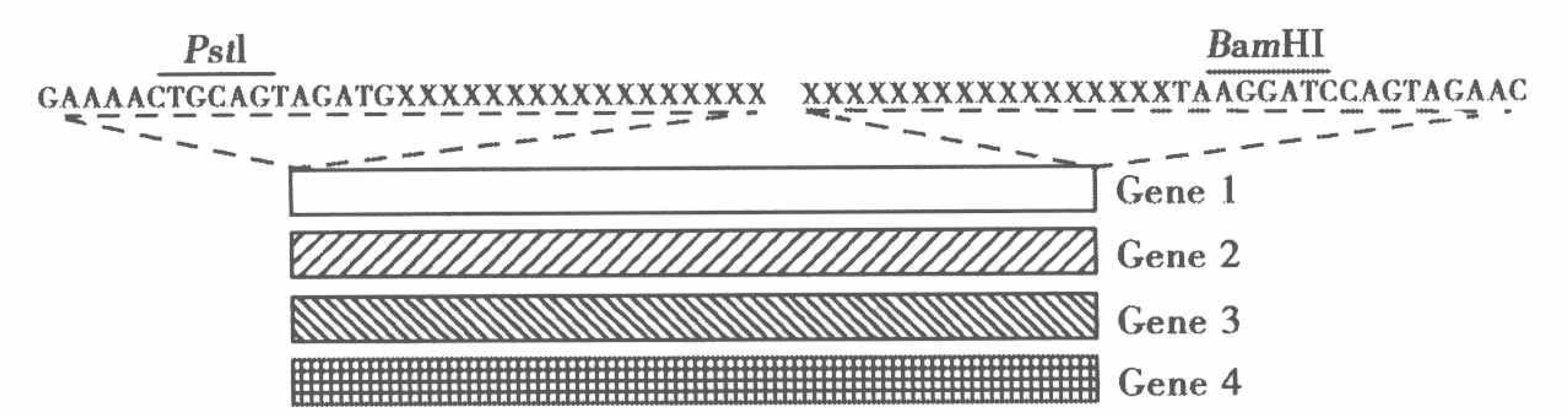
（11）桦木木聚糖底物溶液：0.5g 桦木木聚糖溶于 100ml 的 120mmol/L 通用缓冲液，pH6.5。搅拌 5min 直至溶液均匀，然后高压灭菌。

（12）PAHBAH 储液：分别加入 5ml 的 0.5mol/L 的柠檬酸钠、1.0mol/L 的 Na_2SO_3、0.2mol/L $CaCl_2$ 和 5.0mol/L 的 NaOH 到 25ml 的水中，在加入每种成分前，溶液都要混匀。然后加入 0.76g *p*-羟基苯甲酸肼（*p*-hydroxybenzoic acid hydrazide）(Sigma-Aldrich)，混匀，用水定容到 50ml。

11.3 方　法

我们以 DNA 重组 *D. thermophilum* Rt46B.1 木聚糖酶基因（*xynB*）及其 5 个相关的木聚糖酶基因为例[7]，说明 DOGS 的一般流程。图 11.1 是整个实验的框架图。包括设计引物、引物延伸扩增出基因片段，然后整合这些片段，最后扩增出全长的嵌合基因(注 1)。

A. 设计3′端可与待选基因的N端或C端特异结合，并且5′端结合有通用合适的酶切位点的巢式引物以便能够定向克隆PCR产物。通过聚合酶链反应扩增模板基因



B. 根据候选基因中保守的一个或多个模块，设计出与其互补的简并引物对



C. 利用简并引物以及通用巢式引物，扩增出不同的基因片段

S1　S2　S3　S4

D. 将这些不同的基因片段混合，相邻片段通过重叠区可结合，然后扩增出全长的嵌合基因

S1　S2　S3　S4

E. 酶切和连接上述嵌合基因到合适的载体中，转化到表达菌株，表达出重组蛋白，筛选出满足要求的突变蛋白

图 11.1　DOGS 方法流程

11.3.1　互补的简并引物对可以有效地扩增基因片段并且容易将有重叠的相邻片段连接起来

最常用的分离远缘关系基因的方法，是根据基因中高度保守的序列，设计出简并引物，然后通过 PCR 来分离这些基因。这个方法的难点是，随着引物简并性的增加以适合更多的基因，指导 PCR 反应合成的正确引物数则会降低，而且这些引物在反应最初的几个循环就有可能被用完。此时非特异性的扩增就会在大量的没有参与到目标基因扩增的引物存在下出现，尤其是在错配模板过程中采用不太严格的退火温度时。

为了克服简并寡核苷酸基因改组中简并引物存在的上述问题，Rose 等[13]发明了一种称为共识-简并寡核苷酸混合的方法（CODEHOP）。CODEHOP 引物包含一个相对较短的简并 3′端和一个通用的非简并 5′端。3′端长度减至最少可降低整个简并引物库中的引物数。简并的 3′端与模板的结合可以通过非简并的 5′端来稳定，这样就在不需要增加文库简并度的情况下允许更高的退火温度。虽然 5′端与目标序列在最初的 PCR 循环中可发生潜在的错配，但它们离 3′端延伸位点比较远，因此，错配很少影响聚合酶延伸的起始。基因产物的进一步扩增，通过引物池中引物序列的相似性，被显著增强；这也表明有更多的引物可被使用。DOGS 方法通过改进 CODEHOP，可以更有效地扩增重叠区的基因片段。通过来自不同基因的相邻基因片段的重叠延伸可以导致嵌合基因的生成。合适的引物序列设计，是 DOGS 方法中最重要的一环。

对 CODEHOP 方法的改进，需要很好地设计互补引物对。每个引物有一个非简并的核心，两边是简并的 5′端和 3′端，在这里我们称其为互补简并端（CDE）引物。同 CODEHOP 中的引物一样，简并的 3′端使得 CDE 引物与模板可以特异的结合，而非简并区则可稳定接下来的 PCR 循环。简并的 5′端在 PCR 过程中并不促成 CDE 引物与模板的结合，它主要是用来有效的结合并重叠延伸相互分离的、分别由正向或反向 CDE 引物扩增的 PCR 产物（基因片段）。

每个 CDE 引物非简并的核心区通常是基于相应基因上的编码序列，该基因往往是用来混编的亲代基因。这可以保证嵌合的基因在重组的位点保留亲代基因的序列。

11.3.1.1 设计和使用基因特异的巢式末端引物

（1）设计和合成用来扩增每个需要混编基因的适合的正向和反向引物。每个引物应该包含 17～20 个核苷酸基因特异性的 3′端和包含 17～20 个核苷酸的通用 5′端。我们在通用端引物引入酶切位点，以便定向连接 PCR 产物到 pBSII KS 载体中。

（2）合成两个巢式引物，其序列相对于基因特异的正反向巢式引物通用端。该巢式引物将和 CDE 引物结合使用，以扩增每一基因的首末部分。

（3）用设计好的基因特异性引物扩增出每个基因，一般是从基因组中扩增，或者从其他有该基因的克隆中扩增。

11.3.1.2 CDE 引物特点总结

CDE 引物可以有效地扩增同源关系较小的基因中的片段。CDE 引物的 5′的简并端确保了通过正向和反向互补的 CDE 引物扩增出来的基因片段在接下来的重叠延伸 PCR 步骤中可以有效地退火。而且，由多个 CDE 引物产生的连续基因片段带有适合重叠延伸和 PCR 的互补末端，从而可以产生重组的片段。图 11.2 给出了设计 CDE 引物的一个例子，以便从相关基因中扩增基因片段以及在后续叠加延伸片段中生成嵌合体。

CDE 引物也可以与不含有非简并核心区的互补简并引物结合使用，通过末端互补以及重叠延伸 PCR，从而得到重组的基因片段。将这些来自相关基因的片段混合后进行重叠延伸和 PCR，可以有效地产生嵌合的基因片段。CDE 引物的非简并区可以（但不必须）根据一个亲代基因来设计。最后，来自不同基因的片段可以不等量的混合，这样可以控制不同片段整合产生的新基因（见注 2）。图 11.3 概括了用 CDE 引物扩增和重叠延伸基因片段的流程。

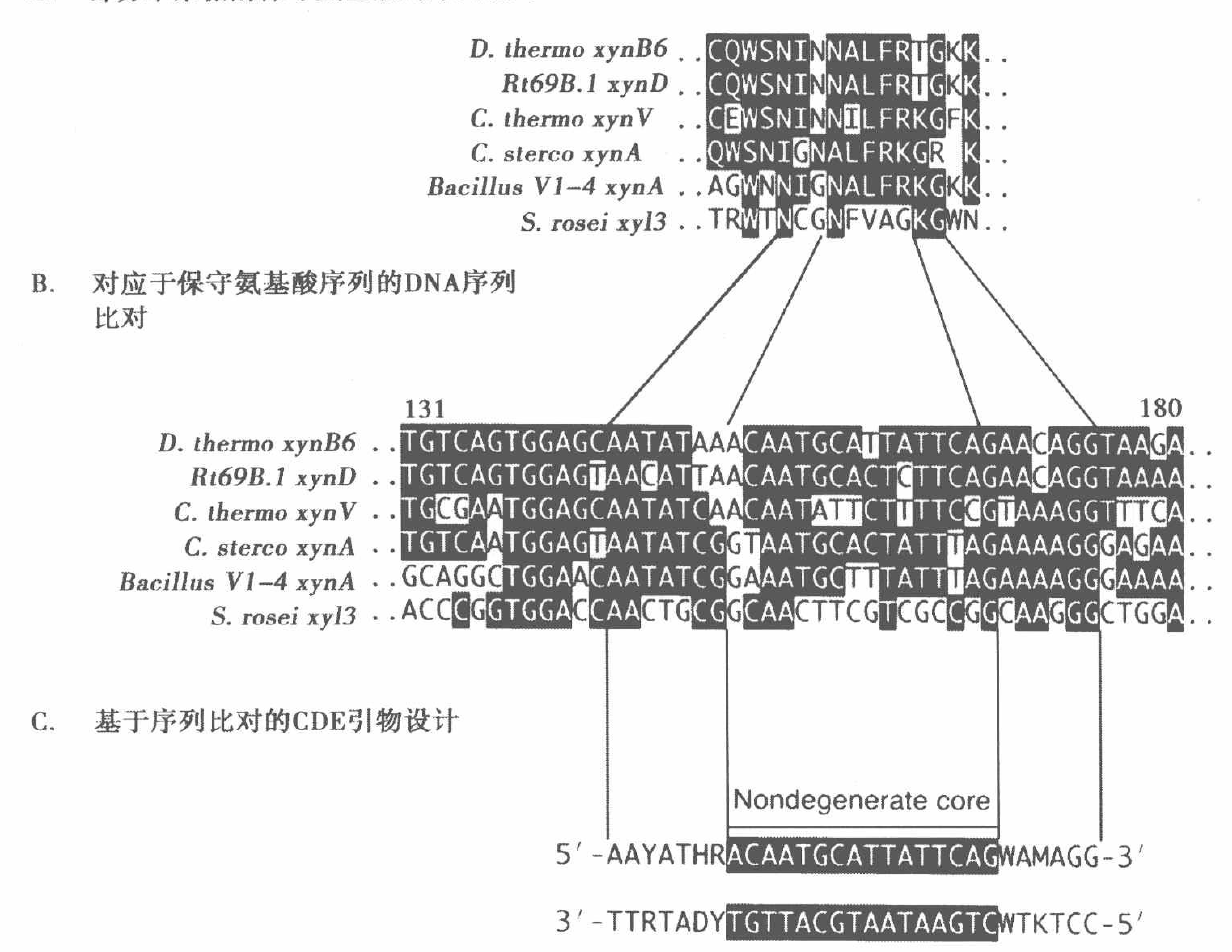


图 11.2　基于保守氨基酸的 DNA 序列进行 CDE 引物的设计例子。A. 部分木聚糖酶保守氨基酸的序列比对。B. 对应于保守氨基酸序列的 DNA 序列比对。C. 基于保守 DNA 序列的 CDE 引物。简并端的设计基于保守的氨基酸和相应的保守的 DNA 序列。非简并核心序列（反向文本）是与亲代基因相匹配的 DNA 序列（在本例中，是 *D. thermophilum xyn* B6 的序列）。

11.3.1.3　设计 CDE 引物以保证有效的基因片段的扩增和相邻片段的连接

(1) 用合适的序列比对软件，如 ClustalX[14]，对相关蛋白质进行序列比对。

(2) 确定保守的氨基酸基序。在我们的例子中，木聚糖酶家族的基因根据其保守的区域可以分成 8 个部分。

(3) 用合适的序列比对软件，如 Tranalign[15]，对基因的核苷酸序列进行比对。

(4) 基于保守的 DNA 序列，设计用以扩增保守位置 DNA 的 CDE 正向和反向引物。然后在适当地加入 5′端和 3′端的通用巢式引物的条件下，扩增出基因片段（图 11.2）。

(5) 用相邻的 CDE 引物和通用的 5′端和 3′端巢式引物，扩增每个基因的每个片段。在这个例子中，PCR 反应的条件是：95℃ 1min，1 个循环；然后 95℃ 30s，35℃ 20s，72℃ 40s，总共 35 个循环；最后 72℃ 5min，1 个循环。在 PCR 反应中使用的是古细菌 DNA 聚合酶，Platinum *Pfx*（见注 3 和注 4）。

(6) 用胶回收每个 PCR 产物。

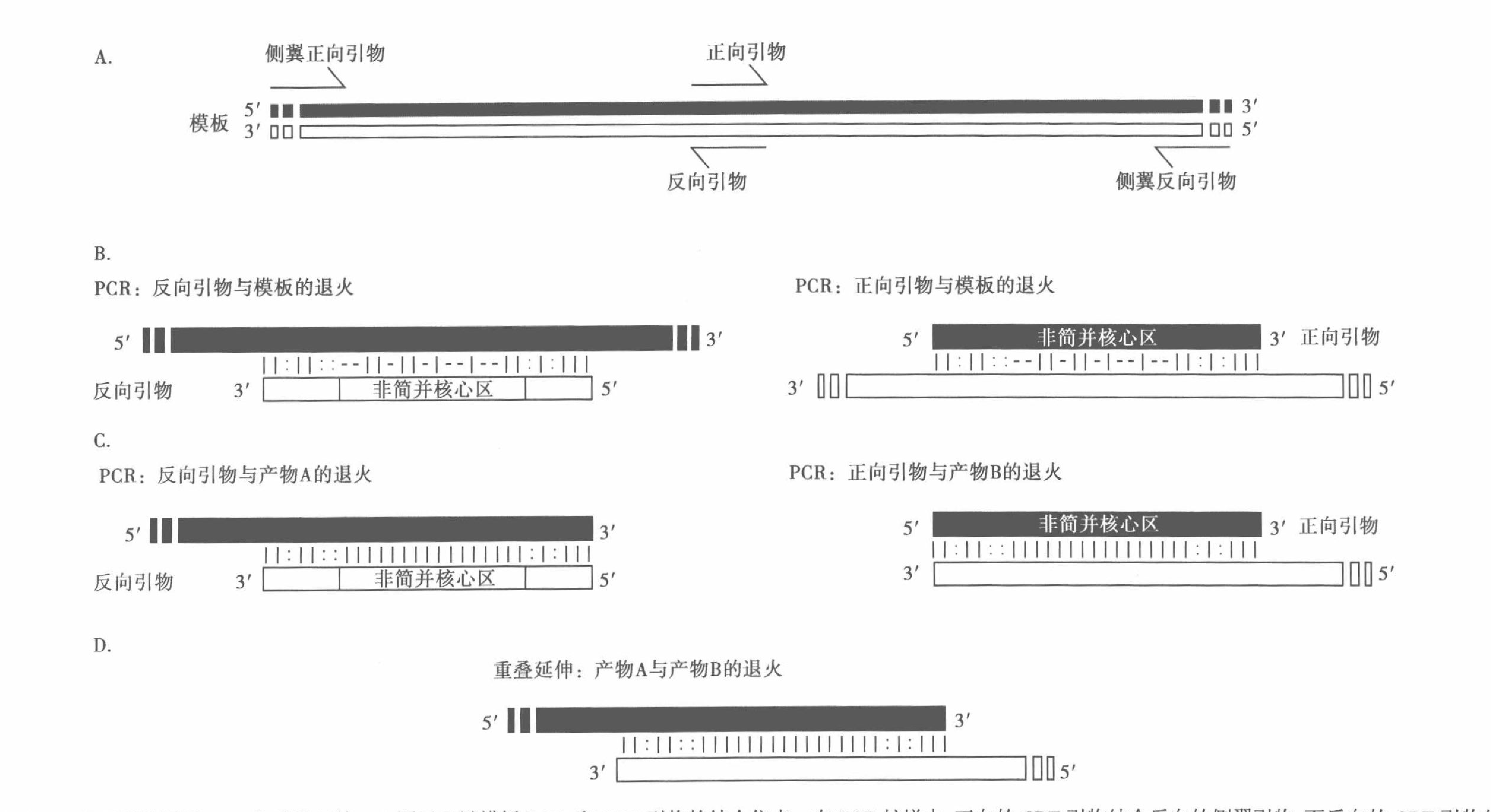


图 11.3　CDE 引物用于 PCR 和重叠延伸。A 图示双链模板 DNA 和 CDE 引物的结合位点。在 PCR 扩增中，正向的 CDE 引物结合反向的侧翼引物，而反向的 CDE 引物结合正向的侧翼引物。B 图示每个 CDE 正向、反向引物与模板 DNA 的结合。(|)表明引物与模板之间核苷酸的正确配对；(：)表示由于引物的简并性潜在的引物与模板核苷酸的配对；(—)表示引物与模板的错配。这里也表明，在 PCR 的第一轮反应中，非简并的核心并不参与引物的结合，引物结合的特异性主要由 3′简并端来实现。C 图示 PCR 扩增反应早期出现的产物。(|)表明引物与模板之间核苷酸的正确配对；(：)表示由于引物的简并性潜在的引物与模板核苷酸的配对。非简并核心起着与模板稳定结合的作用，确保简并引物有效的扩增出基因片段。D 图示表明两个互补的基因片段。这种互补允许高效的聚合酶介导的基因片段的连接，最后产生还有全部 DNA 片段的单链 DNA。如果连接的两个 PCR 片段来自不同的基因，表明重组的基因已经产生。

11.3.1.4　基因片段的重叠延伸

（1）每个基因的片段应当按合适的比例混合，产生基因的嵌合。例如，用 6 个候选的基因 G1～G6，其中 G1 是 *D. thermophilum* Rt46B.1 *xynB* 基因。当决定用 G1 在基因重组中作为主要的亲代基因时，每个基因的 PCR 片段则按照 8.75 倍的 G1 基因相对于等量的其他基因的比例混合，这样，嵌合基因中平均 5/8 的基因片段来自 Rt46B.1 *xynB*（注 5 给出了一个简单的计算公式）。

（2）接下来，用 50～100ng 的混合片段作为模板来进行重叠延伸[16]，应用如下的 PCR 反应体系：95℃ 1min，1 个循环；然后 95℃ 30s，35℃ 20s，72℃ 40s，总共 35 个循环；最后 72℃ 5min，1 个循环。在 PCR 反应中使用的是古细菌 DNA 聚合酶，Platinum *Pfx*（见注 3）。

11.3.1.5　全长嵌合基因的扩增

将重叠延伸的产物（50～100ng，见 11.3.1.4）作为模板，加入通用的 5′端和 3′端巢式引物，然后 PCR 便可得到全长的嵌合重组基因。PCR 的反应条件如下：95℃ 1min，1 个循环；然后 95℃ 30s，50℃ 20s，72℃ 40s，总共 35 个循环；最后 72℃ 5min，1 个循环。在 PCR 反应中用 Platinum *Pfx* 聚合酶。

11.3.1.6　重组基因的克隆

（1）用 *Bam*HI 和 *Hin*dIII 酶切 DOGS 的 PCR 产物。

（2）用 *Bam*HI 和 *Hin*dIII 酶切 pBSII KS-载体，然后用虾碱性磷酸酶处理。

（3）连接 PCR 产物与载体 pBSII KS-。

（4）将连好的载体（载体上带有氨苄抗性）转入到 *E. coli* 的 DH5α 菌株中，然后涂在含有 100μg/ml 氨苄抗性和 5mmol/L IPTG 的 LB 平板上。

（5）挑取单克隆，在新的带有 100μg/ml 氨苄抗性和 5mmol/L IPTG 的 LB 平板上做一备份，然后按 11.3.1.7 节所描述的刚果红覆盖法（Congo Red Overlay）[17]来检测木聚糖酶的表达及活性。

11.3.1.7　刚果红筛选法

（1）将 4ml 冷却至大约 50℃的覆盖液加入到长有克隆的平板上。平板应该先在 37℃预热以确保均匀的覆盖层的形成。

（2）当覆盖层形成后，将平盘倒置在一个封口袋内。在 70℃密封放置 3h。

（3）然后取出平板，去掉袋子，放置室温冷却。

（4）在每个平板上加入 5ml 刚果红溶液，使其能全部覆盖覆盖层。然后放置 5～10min。

（5）倒出剩余的刚果红溶液，然后将平板浸润在脱色液里脱色。

（6）未染上色的克隆是含有木聚糖酶基因阳性克隆。

图 11.4 为用 DOGS 方法从 6 种木聚糖酶的基因中产生重组木聚糖酶的示例（见注 6）。

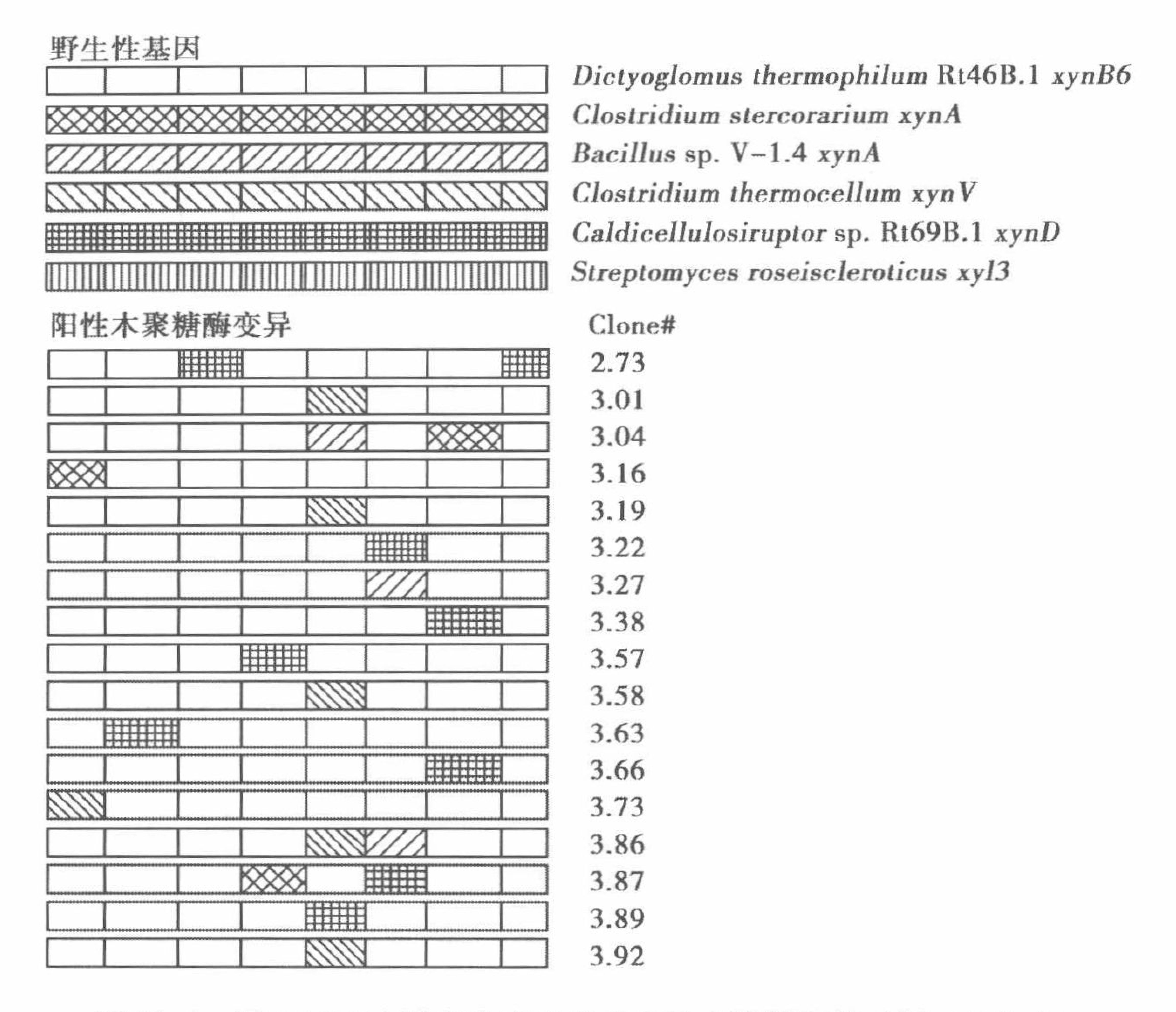


图 11.4 用 DOGS 方法产生有功能的重组木聚糖酶的示例（见注 6）。

11.3.1.8 木聚糖酶活性测定（PAHBAH 方法）

使用蛋白质抽提试剂 BPER II，在全细胞中提取重组的木聚糖酶用来酶活分析。木聚糖酶的活性分析是用 Lever 法[18]，以桦木木聚糖为底物测定的。终体积为 0.03ml 的标准的反应混合液包括：120mmol/L 的通用缓冲液（pH6.5）[19]、0.5%（*m*/*V*）木聚糖、还原酶。反应混合液在 60℃放置 20min。

1）细胞裂解和酶的提取

（1）将鉴定为阳性转化株的在 2ml 含有 100μg/ml 氨苄和 IPTG 的 LB 中过夜培养。

（2）取出 1.5ml 的过夜培养的 LB，11 000*g* 离心 30s，去掉上清液。

（3）通过振荡悬浮上述细胞，然后加入 150μl 的 BPER II 溶液。

（4）振荡混匀 1min，裂解细胞，11 000*g* 离心 1min，去掉细胞碎片。

（5）将上清液转移到新的管中，然后在 4℃储存。

2）木聚糖酶活性试验

（1）在 0.2ml PCR 管中，将 5μl 适当稀释的酶提取液加入到桦木木聚糖底物溶液中至终体积为 50μl。

（2）将 PCR 管放在顶部能够加热的 PCR 仪中，在合适的温度下温育 10～30min。

（3）将 PCR 管放在冰水中，加入 PAHBAH 储液 100μl 终止酶活反应，混匀。

（4）将上述混合液在 PCR 仪中 99℃加热 5min 进行变色反应。确保可加热的盖子置于关闭的位置。然后立即冷却到 4℃，以停止颜色的变化。

（5）将管中的溶液混匀，然后转移 100μl 到平底的微孔板中。用具有合适的可读板的光谱仪检测在 A_{420} nm 的吸收值。

11.4　注

1. DOGS方法使同源关系较小的基因发生高频率的重组成为可能，而且不需要费时费力的筛选鉴定。CDE引物可以从整体同源性不高的基因中有效地扩增出用于重组的基因片段。尽管在我们给出的例子中，是用不同物种的木聚糖酶基因进行混编，这一方法也可以被推广到对单一基因的突变库进行混编。可以用最有希望的突变体来替代家族中不同的基因。利用CDE引物，将不同的突变体用PCR得到，然后按一定的比例混合这些突变体，按照DOGS所描述的方法，进行基因改组，可以产生有害的或中性的突变。

2. 某个片段重组的频率可以根据改变该片段加入量的比例来控制，使其发生重组的频率按要求升高或降低（见注4）。相应地，DOGS方法也可以相对容易的进行结构域的交换，这样就可以替代较早的利用适当的内切酶来进行的交换了。将错掺突变PCR整合到DOGS方法中，可以产生多样性的重组。在这方面，DOGS可以仅对基因的某一部分进行突变，从而可以帮助编码基因更好地调控。在这里，可以改变CDE引物，使其只对突变的基因片段有简并性。这样，研究者就可以通过在适合的阶段引入保真性缺失的DNA聚合酶来控制嵌合基因片段上的特性。

3. 高保真的Platinum *Pfx*聚合酶被用在所有的PCR反应中，以降低错掺突变的发生率。应尽量避免*Taq*酶的使用，因为该酶会使PCR产物的3′端多个A，从而使接下来的重叠延伸PCR片段间的退火受到干扰。

4. 由于在简并引物的设计中会有错配的设计，所以即使是用高保真DNA聚合酶，PCR也会使核苷酸序列发生改变。如果在扩增和引物延伸步骤中选用保真性不高的酶，则会引入更多的错掺突变。

5. 为了获得需要的嵌合基因，可以用公式 $r=(gx-x)/(s-x)$ 来计算亲代基因片段与其他基因片段的混合比例 r。其中 g 是用来混编的基因的个数；x 是每个嵌合基因要求的平均亲代基因的片段数；s 是每个基因的片段数。例如，我们想重组6个基因，每个基因分成8个片段；平均每个嵌合基因要包括6个亲代片段，这样 $g=6$，$s=8$，$x=6$。用上述公式计算，r 为15。因此，亲代片段应该以15∶1∶1∶1∶1∶1的比例与来自其他5个基因的片段混合，以期获得平均含有6个亲代片段的嵌合基因。

6. 显而易见，利用DOGS方法，再结合其他基因混编的方法，则可以产生具备新特性的重组蛋白来。

致谢　这项工作由澳大利亚科研委员会、麦夸里大学研究基金会和新西兰公益科学基金支持的。

参考文献

1. Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. C. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* 14, 315-319.

2. Stemmer, W. P. C. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389-391.

3. Stemmer, W. P. C. (1994) DNA shuffling by random fragmentation and reassem-bly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. USA* 91, 10747-10751.

4. Schmidt-Dannert, C., Umeno, D., and Arnold, F. H. (2000) Molecular breeding of carotenoid biosynthetic pathways. *Nature Biotechnol.* 18, 750-753.

5. Kikuchi, M., Ohnishi, K., and Harayama, S. (1999) Novel family shuffling methods for the *in vitro* evolu-

tion of enzymes. *Gene* 236, 159-167.

6. Morris, D. D., Gibbs, D. D., Chin, C. W., et al. (1998) Cloning of the xynB gene from *Dictyoglomus thermophilum* strain Rt46B. 1 and action of the gene-product on kraft pulp. *Appl. Environ. Microbiol.* 64, 1759-1765.
7. Gibbs, M. D., Nevalainen, K. M. H., and Bergquist, P. L. (2001) Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. *Gene* 271, 13-20.
8. Sakka, K., Kojima, Y., Kondo, T., Karita, S., Ohmiya, K., and Shimada, K. (1993) Nucleotide sequence of the *Clostridium stercorarium xynA* gene encoding xylanase A: identification of catalytic and cellulose binding domains. *Biosci. Biotechnol. Biochem.* 57, 273-277.
9. Yang, V. W., Zhuang, Z., Elegir, G., and Jeffries, T. W. (1995) Alkaline-active xylanase produced by an alkaliphilic *Bacillus* sp. isolated from kraft pulp. *J. Ind. Microbiol.* 15, 434-441.
10. Morris, D. D., Gibbs, M. D., Ford, M., Thomas, J., and Bergquist, P. L. (1999) Family10 and 11 xylanase genes from Caldicellulosiruptor isolate Rt69B. 1. *Extremophiles* 3, 103-111.
11. Fernandes, A. C., Fontes, C. M., Gilbert, H. J., Hazlewood, G. P., Fernandes, T. H., and Ferreira, L. M. (1999) Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochem. J.* 342, 105-110.
12. Elegir, G., Szakacs, G., and Jeffries, T. W. (1994) Purification, characterization and substrate specificity of multiple xylanases from *Streptomyces* sp. strain B-12-2. *Appl. Environ. Microbiol.* 60, 2609-2615.
13. Rose, T. M., Schultz, E. R., Henikoff, J. G., Pietrokovski, S., McCallum, C. M., and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* 26, 1628-1635.
14. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* 24, 4876-4882.
15. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276-277.
16. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77, 51-59.
17. Teather, R. M. and Wood, P. J. (1982) Use of Congo Red polysaccharide interaction in enumeration and characterisation of cellulolytic bacteria from bovine rumen. *Appl. Environ. Microbiol.* 43, 777-780.
18. Lever, M. (1973) Colorimetric and fluorometric carbohydrate determination with *p*-hydroxybenzoic acid hydrazide. *Biochem. Med.* 7, 274-281.
19. Britton, H. T. S. and Robinson, R. A. (1931) Universal buffer solutions and the dissociation constant of veronal. *J. Chem. Soc.* 1, 1456-1462.

12　M13 噬菌体衣壳蛋白改造在改良噬菌体展示技术中的应用

Sachdev S. Sidhu，Birte K. Feld，and Gregory A. Weiss

概要　本章描述了改进融合蛋白在 M13 噬菌体颗粒表面展示水平的一个方法。向 M13 衣壳锚定蛋白引入突变后，蛋白质展示水平约能增加两个数量级。这里讲述改进蛋白质展示水平的噬菌体展示库的设计、构建以及筛选方法。

关键词　噬菌体展示；蛋白质工程；组合突变；M13 噬菌体；主要衣壳蛋白；病毒进化

12.1　引　言

噬菌体展示是一项改造结合靶分子多肽的强大技术[1～3]。蛋白质与噬菌体衣壳蛋白形成融合蛋白，可以在噬菌体表面表达，而相应的编码基因则包裹于噬菌体颗粒内[4]。包裹的 DNA 通过突变扩增构建噬菌体展示多肽库，为筛选各种变异体做准备。通过固定化靶分子体外结合实验，高亲和力的蛋白质可以从蛋白质展示库中筛选出来。随后，选定蛋白质的序列可以通过 DNA 测序推断出来。

靶蛋白可以通过构建噬菌粒载体与主要衣壳蛋白［蛋白质-8（P8）］或次要衣壳蛋白［蛋白质-3（P3）］形成的融合蛋白在 M13 噬菌体表面以低拷贝方式展示，这需要在野生型（wt）P8 和 P3 辅助噬菌体帮助的条件下进行[5]。由于融合蛋白的展示水平随蛋白质序列和长度的不同而不同，多价蛋白与 P8 融合进行展示难以实现[6]。例如，如果 P8 的每一个拷贝都与展示多肽形成融合蛋白（噬菌体系统，而不是噬菌粒系统），融合多肽长度超过 6 个氨基酸残基，噬菌体一般就会不稳定[7]。即使在噬菌粒系统中，蛋白质展示也高度依赖于蛋白质的大小和性质[8]。尽管用 P3 或 P8 进行单价噬菌体展示使许多不同蛋白质亲和力从中等提高到很高[9]，利用 P8 进行多价噬菌体展示从初级库中选择分子质量大的低亲和力受体蛋白尚未实现。

这里，我们描述一种通过突变 P8 锚定残基改进蛋白质展示的方法。利用这一方法，两个不同的蛋白质［(链霉亲和素和人生长激素（hGH)］表达水平几乎增加了 100 倍，并且我们发现这一技术也适用于其他蛋白质。这一戏剧性的展示改进很可能由于融合蛋白在噬菌体表面定位更加恰当所致。理论上，高拷贝展示的应用应该可能使结合弱的蛋白质从初级天然库中选择出来，对于小肽这一方法已经成功可行。同样重要的是这一方法扩展了噬菌体展示研究蛋白质的范围，许多曾被认为难以展示的蛋白质用这一方法实现了展示。因此，这里描述的方法增加了噬菌体展示鉴定新蛋白质或修饰蛋白质功能的潜力。

12.2 材　　料

12.2.1 P8突变库的构建

12.2.1.1 制备含有尿嘧啶的单链DNA模板

(1) 2YT 培养基：10g 酵母抽提物、16g 菌用胰蛋白胨和 5g 氯化钠；加水到 1L，用氢氧化钠调节 pH 到 7.0；高压灭菌。

(2) 2YT/carb/cmp 培养基：2YT、50μg/ml 羧苄青霉素（carbenicillin，carb）和 5μg/ml 氯霉素（chloramphenicol，cmp）。

(3) 2YT/carb/kan/uridine 培养基：2YT、50μg/ml 羧苄青霉素、25μg/ml 卡那霉素和 0.25μg/ml 尿嘧啶。

(4) 羧苄青霉素：5mg/ml，羧苄青霉素水溶液，超滤灭菌。

(5) 氯霉素：50mg/ml，氯霉素乙醇溶液。

(6) 大肠杆菌 CJ236（NEB：New England Biolabs，Beverly，美国麻省）。

(7) 卡那霉素：5mg/ml，卡那霉素水溶液，超滤灭菌。

(8) 生理磷酸缓冲液（PBS）：137mmol/L 氯化钠、3mmol/L 氯化钾、8mmol/L 磷酸氢二钠以及 1.5mmol/L 磷酸二氢钾；用盐酸（HCl）调节 pH 到 7.2；高压灭菌。

(9) 聚乙二醇（polyethylene glycol，PEG）PEG/NaCl：20% PEG-8000（*m/V*）和 2.5mol/L 氯化钠；高压灭菌。

(10) QIAprep SpinM13 试剂盒（Qiagen，Valencia，美国加州）。

(11) M13KO7 辅助噬菌体（NEB，Beverly，美国麻省）。

(12) Tris-acetate-EDTA（TAE）缓冲液：40mmol/L Tris-acetate 和 1mmol/L EDTA；调节 pH 到 8.0；高压灭菌。

(13) TAE/琼脂糖胶：TAE 缓冲液、1% 琼脂糖（*m/V*）以及 1∶5000（*V/V*）10%溴化乙锭（EB）。

(14) 尿嘧啶：25mg/ml，尿嘧啶水溶液，超滤灭菌。

12.2.1.2 体外合成异质共价闭环双链DNA

(1) 100mmol/L 二硫苏糖醇（DTT）。

(2) 25mmol/L dNTP：含有 dATP、dCTP、dGTP 和 dTTP 各 25mmol/L 的溶液（Amersham-Pharmacia，Piscataway，美国新泽西州）。

(3) 10mmol/L ATP（Amersham-Pharmacia）。

(4) 10× Tris-Mg（TM）缓冲液：0.1mol/L $MgCl_2$ 以及 0.5mol/L Tris-HCl，pH7.5。

(5) QIAquick Gel Extraction 试剂盒（Qiagen，Valencia，美国加州）。

(6) T4 多聚核苷酸激酶（NEB：New England Biolabs）。

(7) T4 DNA 连接酶（Invitrogen，Carlsbad，美国加州）。

(8) TAE 琼脂糖凝胶（见 12.2.1.1）。

(9) USP (United States Pharmacopeia) 超纯水 (Braun Medical Inc.，Irvine，美国加州)。

12.2.1.3 大肠杆菌电穿孔及噬菌体扩增

(1) 2YT/carb 培养基：2YT 及 50μg/ml 羧苄青霉素。

(2) 羧苄青霉素 (carbenicillin，carb) (见 12.2.1.1)。

(3) 电转化感受态 *E. coli* SS320 (见 12.3.2.4)。

(4) 0.2cm 间隙电转化样品杯 (BTX，圣地亚哥，美国加州)。

(5) ECM-600 电转化仪 (BTX)。

(6) 卡那霉素 (kanamycin，kan) (见 12.2.1.1)。

(7) Luria-Bertani (LB) /carb 培养皿：LB 琼脂及 50μg/ml 羧苄青霉素。

(8) PBS (见 12.2.1.1)。

(9) PEG/NaCl (见 12.2.1.1)。

(10) SOC 培养基：5g 细菌用酵母抽提物、20g 细菌用胰蛋白胨和 0.5g 氯化钠及 0.2g 氯化钾；加水到大约 1L，用氢氧化钠调节 pH 到 7.0；高压灭菌；然后加入 5ml 2.0mol/L 高压灭菌过的氯化镁溶液及 20ml 超滤灭菌过的 1.0mol/L 葡萄糖溶液；并将最终体积调至 1L。

(11) M13KO7 辅助噬菌体 (见 12.2.1.1)。

12.2.1.4 制备 *E. coli SS320* 电穿孔法感受态细胞

(1) 1.0mmol/L HEPES，pH7.4：将 4.0ml 1.0mol/L 的 HEPES，pH7.4 加入 4.0L 的 USP (United States Pharmacopeia) 超纯水，超滤灭菌。

(2) 10% (*V*/*V*) 的超纯甘油：100ml 超纯甘油加入 900ml USP 超纯水中，超滤灭菌。

(3) 2YT/tet 培养基：2YT 及 5μg/ml 四环素。

(4) 磁性搅拌棒 (约 5cm 长)，放入乙醇中浸泡。

(5) Superbroth/tet 培养基：24g 细菌用酵母抽提物、12g 细菌用胰蛋白胨和 5ml 甘油；加水到 900ml；高压灭菌；加入 100ml 高压灭菌的 0.17mol/L 磷酸二氢钾和 0.72mol/L 的磷酸氢二钾混合液，最后加入 5μg/ml 四环素。

(6) 四环素 (tetracycline)：5mg/ml 四环素水溶液，超滤灭菌。

(7) 超纯甘油 (Invitrogen)。

(8) USP (United States Pharmacopeia) 超纯水 (见 12.2.1.2)。

12.2.2 选择和分析增加融合蛋白展示水平的 P8 变异体

12.2.2.1 从 hGH-P8 库中选择噬菌体

(1) 0.2%牛血清白蛋白 (BSA) 的 PBS 溶液。

(2) 100mmol/L HCl。

(3) 1.0mol/L Tris 碱。

(4) 96 孔 maxisorp 免疫板 (NUNC，Roskilde，丹麦)。

(5) 大肠杆菌 *E. coli* XL-1 Blue 菌株 (Stratagene，La Jolla，美国加州)。

（6）PBS-T 缓冲液：PBS 及 0.05%Tween-20。

（7）PBS-T-BSA 缓冲液：PBS，0.05%Tween-20 以及 0.2% BSA。

（8）2YT/carb/kan 培养基：2YT、50μg/ml 羧苄青霉素以及 25μg/ml 卡那霉素。

12.2.2.2 噬菌体酶联免疫筛选法测定 hGH 的展示水平

（1）1.0mol/L H_3PO_4。

（2）2YT/carb/kan 培养基（见 12.2.2.1）。

（3）3,3′,5,5′-四甲基联苯胺（TMB）/H_2O_2 过氧化物酶底物：（Kirkegaard&Perry Laboratories Inc.，Gaithersburg，MD）。

（4）96 孔 maxisorp 免疫板（见 12.2.2.1）。

（5）羧苄青霉素（carbenicillin，carb）（见 12.2.1.1）。

（6）大肠杆菌 *E. coli* XL-1 Blue 菌株（见 12.2.2.1）。

（7）辣根过氧化物酶（Horseradish peroxidase）/抗 M13 抗体络合物（Amersham-Pharmacia）。

（8）卡那霉素（见 12.2.1.1）。

（9）LB/tet 培养皿：LB 琼脂及 5μg/ml 四环素。

（10）M13KO7 helper 噬菌体（见 12.2.1.1）。

（11）PBS（见 12.2.1.1）。

（12）PBS-T 缓冲液（见 12.2.2.1）。

（13）PBS-T-BSA 缓冲液（见 12.2.2.1）。

（14）PEG/NaCl（见 12.2.1.1）。

（15）四环素（见 12.2.1.4）。

12.3 方　　法

下述方法描述了 P8 库的设计（见 12.3.1）、构建（见 12.3.2）、选择（见 12.3.3.1）和筛选（见 12.3.3.2），依此来鉴定能改进蛋白质表达水平的突变体。我们描述了利用前述噬菌粒——pS1607 改进 hGH 表达水平的方案。这一方法也适用于任何其他靶蛋白。唯一的改变就是用表达靶蛋白的噬菌体粒取代 pS1607 以及用高亲和性的靶蛋白配体取代 hGH 结合蛋白。

12.3.1 库的设计

我们前面已经提及在噬菌粒表达系统中 P8 的 N 端部分极度耐受突变，其中有些突变能增加异源融合蛋白的展示水平[10,11]。随后，我们发现 P8 的 N 端只有 6 个野生型残基侧链（Ala7、Ala9、Ala10、Phe11、Leu14 和 Ala18）对 P8 有效包装到噬菌体外壳是必需的[10,12]。这些侧链形成一个紧密的疏水抗原决定簇（图 12.1），在噬菌体组装过程中起到关键作用。环绕这一决定簇的残基突变能增加包装效率，从而增加异源蛋白的展示水平[10]。基于这些研究，我们鉴定出了 7 个位点（Pro6、Lys8、Asn12、Ser13、Gln15、Ala16 和 Ser17），这些位点的突变最可能改进蛋白质展示水平（图

12.1)。蛋白质内 7 个位点的完全随机化能导致接近 10^9 的独特氨基酸组合，这一多样性能够被这里所描述的制备库的约 10^{10} 的多样性所覆盖。

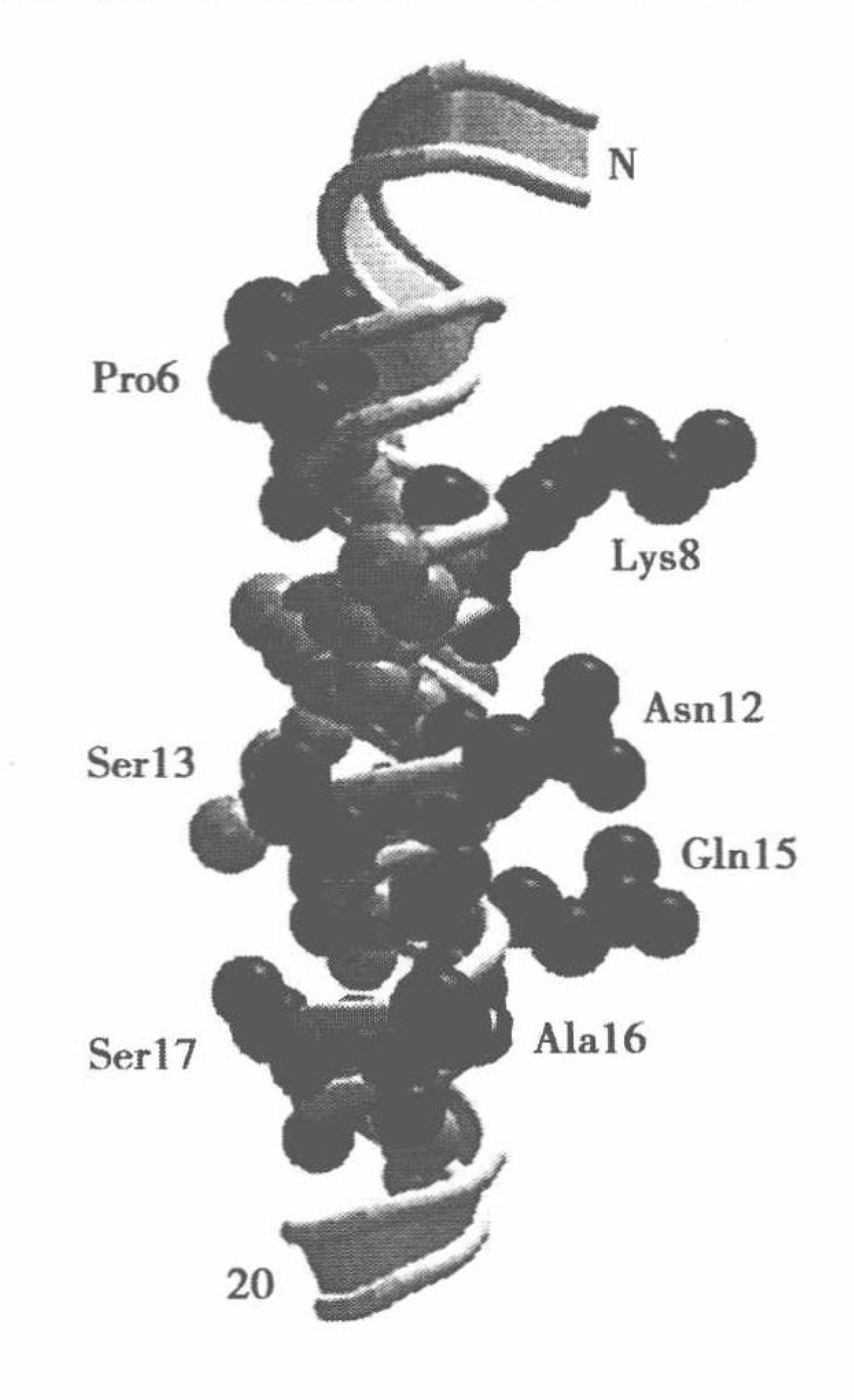


图 12.1　丝状噬菌体主要衣壳蛋白 P8。P8 主链的前 20 个残基显示为螺旋飘带图。有效掺入噬菌体衣壳所需的野生型侧链残基显示为浅灰色（Ala7、Ala9、Ala10、Phe11、Leu14 以及 Ala18)。为增加外源融合蛋白展示水平的突变靶残基显示为黑色并且标出。这里显示的 P8 来自 Ff 噬菌体，其只和 M13 的 P8 相差一个位置（M13 P8 中的 Asn 12 在 Ff P8 中是 Asp)。原子坐标来自于布鲁克海文蛋白质数据库（PDB 编码 1IFJ)，利用 Visual Molecular Dynamics[17] 和 RASTER-3D 软件[18] 显示。

12.3.2　库的构建

P8 变异库用前述寡核苷酸诱导突变法的优化版本[2] 来构建[10]。首先，一条突变的寡核苷酸（序列：GCCGAGGGTGACGATTAAGCATAAGCGGCCTTTTAATAACT-GTAATAATAAGCGACCGAATATATC）用于在随机化位点引入终止密码子；12.3.2.2 节中提供的寡核苷酸诱导的位点特异的突变方案可以用于小规模突变来引入终止密码子（见注 1)。终止密码子的引入消除了野生型蛋白的表达，所以“终止模板”噬菌粒可以作为模板用来建库。含尿嘧啶的单链 DNA 终止模板（从 *E. coli dut-/ung-* 宿主中纯化的）和突变寡核苷酸（序列：GCCGAGGGTGACGATNNKGCATA-AGCGGCCTTTNNKNNKCTGNNKNNKNNKGCCGACCGAATATATC）进行退火，这样就能用编码 20 种氨基酸的 NNK（N＝A/G/C/T，各 25%；K＝G/T，各 50%）取代终止密码子。突变寡核苷酸链作为引物合成一条互补的 DNA 链，最终形成一个共价闭环双链异质 DNA（CCC-dsDNA)。建库完成后，CCC-dsDNA 通过电转化法引入 *E. coli* dut^+/ung^+ 宿主中，在宿主中错配按照野生序列或突变序列进行修复。在 ung^+ 菌株中，dU 模板链倾向于失活，而合成的突变链被增殖，这样就实现了有效突变（大于 50%)。用随机化位点均为终止密码子的序列做模板保证了只有完全突变的克隆才含有可读框，才能在噬菌体表面展示。与辅助噬菌体一起转化宿主大肠杆菌，库里的各成员就能包装进噬菌体颗粒。

12.3.2.1　纯化 dU-ssDNA 模板

突变效率依赖模板的纯度，因此高纯度 dU-ssDNA 的应用对库的成功构建至关重要。我们用 Qiagen QIAprep Spin M13 Kit 来纯化 dU-ssDNA，下面就是 Qiagen 方案

的一个修改版。对一个中等拷贝的噬菌粒而言（如 pS1607，其包含 pBR322 的骨架），这个方案至少能获得 20μg 的 dU-ssDNA，这足以满足建库的需要（见注 2）。

（1）从新鲜的 LB/抗菌素板上挑出一个含有某噬菌粒的 *E. coli* CJ236（或者其他 dut^-/ung^-）菌株的单克隆放入 1ml 2YT 加有 M13KO7 helper 噬菌体（10^{10} pfu/ml）和相应抗菌素的培养基中以保持宿主 F′附加体和噬菌粒。例如，2YT/carb/cmp 培养基中含有羧苄青霉素来选择带有 β-内酰胺酶基因的噬菌粒，而氯霉素用来选择 CJ236 F′附加体。37℃及 200r/min 下摇动 2h 并且加入卡那霉素（25μg/ml）来选择被 M13KO7 共转染的带有卡那霉素抗性基因的克隆。在 37℃及 200r/min 下摇动 6h 后，转移培养到 30ml 的 2YT/carb/kan/uridine 培养基。再在 37℃及 200r/min 下摇动过夜。

（2）在 4℃ 27 000*g* 下（15 000r/min 在 Sorvall SS-34 转子中）离心 10min。转移上清液到含有 1/5 体积 PEG/NaCl 的新试管中，室温孵育 5min。在 4℃ 12 000*g* 下（10 000r/min 在 Sorvall SS-34 转子中）离心 10min。移走上清液，在 2000*g*（4000r/min）下短暂离心，吸出残余上清液。

（3）在 0.5ml 的 PBS 中重新悬浮起噬菌体沉淀小团，转移到一个新的 EP 离心管中。在桌式小离心机中用 14 000*g* 离心 5min，转移上清液到新的 EP 离心管中。

（4）加入 7.0μl MP 缓冲液（Qiagen），混匀。室温孵育至少 2min。

（5）在 2ml 小离心试管的 QIAprep spin column（离心层析柱，Qiagen）中加入上述样品。在桌式小离心机中 6000*g* 离心 30s。遗弃流出液。噬菌体颗粒仍然结合在层析柱介质中。

（6）在柱中加入 0.7ml 的 MLB 缓冲液（Qiagen）。6000*g* 离心 30s，弃流出液。

（7）在柱中再加入 0.7ml 的 MLB 缓冲液。室温孵育至少 1min。6000*g* 离心 30s。弃流出液。噬菌体 DNA 与壳蛋白分离，仍然吸附在层析柱介质中。

（8）加入 0.7ml 的 PE 缓冲液（Qiagen）。6000*g* 离心 30s，弃流出液。

（9）重复步骤（8），去除残余的蛋白质及盐。

（10）6000*g* 离心 30s，将小层析柱转移到一个新的 1.5ml EP 离心试管中。

（11）加入 100μl 的 EB 缓冲液（Qiagen：10mmol/L Tris-HCl，pH8.5）到层析柱膜的中心部位。室温孵育 10min，然后 6000*g* 离心 30s。保存流出液，其中包含纯化的dU-ssDNA。

（12）分析上述 DNA，用 1.0μl DNA 溶液进行 TAE 琼脂糖凝胶电泳。结果 DNA 应该是明显的单一条带，但是也经常可以观察到一些较低电泳迁移率的弱带（见图 12.2 中泳道 2）。这些弱带很可能是由于 ssDNA 的二级结构所致。

（13）利用 260nm 的吸收值（$A_{260}=1.0$ 相应于 33ng/μl 单链 DNA）测定 DNA 浓度。典型 DNA 浓度范围应该为 200～500ng/μl。

12.3.2.2 体外合成异源双链 CCC-dsDNA

用 dU-ssDNA 作为模板，经过三步程序就能把突变寡核苷酸包装到异源 CCC-dsDNA 内。这里描述的方案是前述方法的改进和规模扩大化的版本[13]。寡核苷酸先进行 5′磷酸化，然后和 dU-ssDNA 模板退火。寡核苷酸链通过延伸和连接形成异源 CCC-dsDNA（图 12.2 中泳道 3），然后再进行纯化和脱盐。这一方案能获得大约 20μg 高纯度、低电导率的 CCC-dsDNA。这足以构建容量超过 10^{10} 的库（见注 3）。

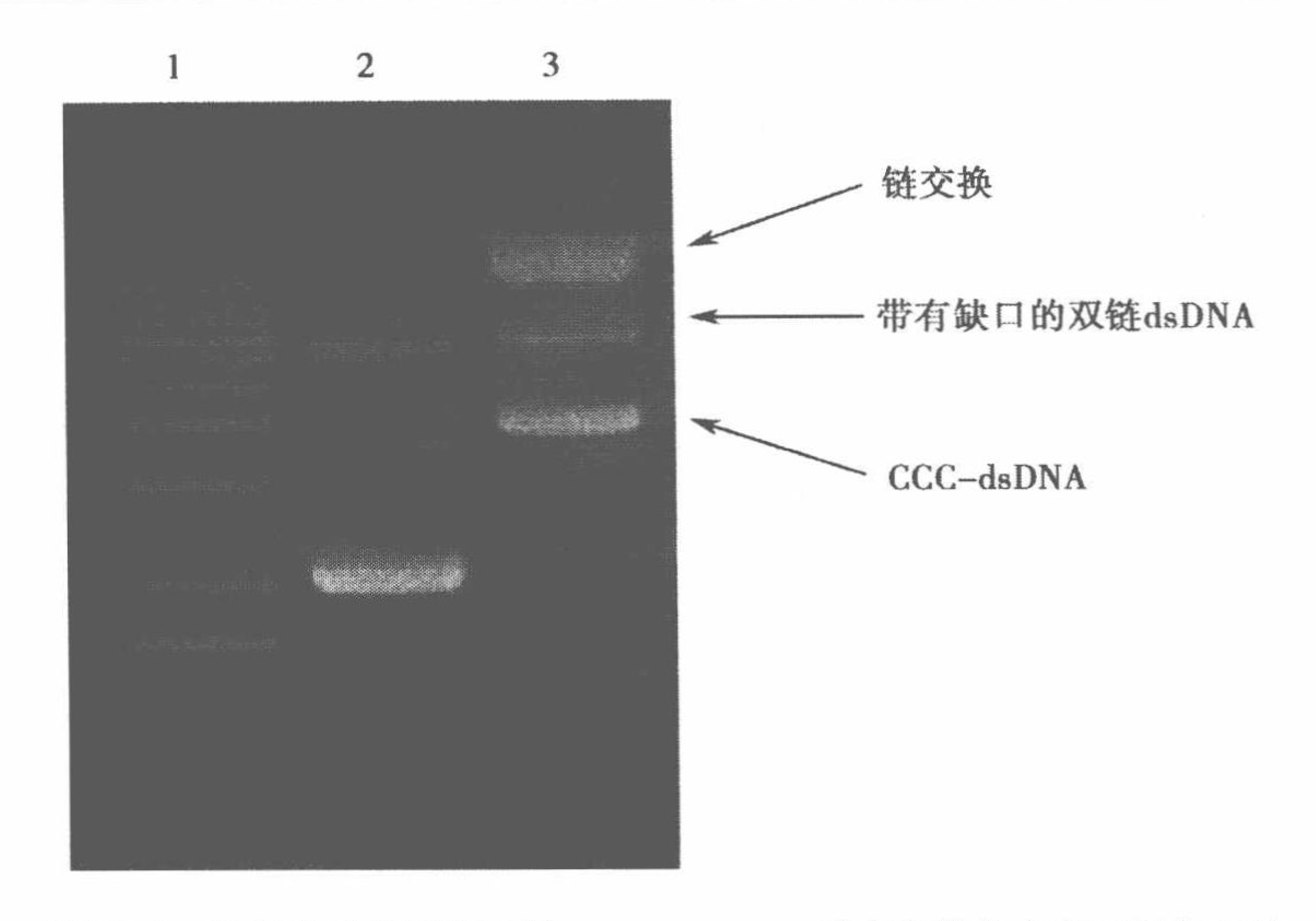


图 12.2　体外合成的异源双链 CCC-dsDNA。反应产物由含有可显色双链 DNA 的 EB 染色剂的 1.0% TAE/琼脂胶电泳显示。泳道 1：1kb DNA 标记（Gibco BRL）；泳道 2：dU-ssDNA 模板；泳道 3：来自 12.3.2.2 中的反应产物。下部条带是正确延伸及连接好的 CCC-dsDNA，中间条带是带有缺口的双链 DNA，上部条带是链交换的双链 DNA。

1）用 T4 多聚核苷酸激酶磷酸化寡核苷酸

（1）在 1.5ml EP 离心管中混合加入 0.6μg 突变寡核苷酸、2.0μl 的 10×TM 缓冲液、2.0μl 的 10mmol/L ATP 和 1.0μl 的 100mmol/L DTT。加水到总体积为 20μl。

（2）往上述体系中加入 20U 的 T4 聚核苷酸激酶。37℃孵育 1h（见注 4）。

2）寡核苷酸和模板一起退火

（1）在 20μl 磷酸化反应混合物中加入 20μg 的 dU-ssDNA 模板（来自 12.3.2.1）、25μl 的 10×TM 缓冲液，加水到总体积为 250μl。假设寡核苷酸与模板长度比是 1∶100，上述 DNA 的量给出寡核苷酸与模板摩尔比为 3∶1。

（2）90℃孵育 3min、50℃孵育 3min、20℃孵育 5min（见注 5）。

3）合成 CCC-dsDNA

（1）在已经退火的寡核苷酸与模板混合物中加入 10μl 10mmol/L ATP、10μl 25mmol/L 的 dNTP、15μl 100mmol/L DTT、30 Weiss U 的 T4 DNA 连接酶以及 30U 的 T7 DNA 聚合酶。

（2）20℃孵育过夜。

（3）利用 Qiagen QIAquick DNA 纯化试剂盒，对上述 DNA 进行亲和纯化及脱盐。加入 1.0ml 的 QG（Qiagen）缓冲液混合。

（4）将上述样品放入两个置于 2ml 离心试管中的 QIAquick spin 离心层析柱。在桌式小离心机中用 14 000*g* 离心 1min，去除流过液。

（5）在每个柱子中加入 750μl 的 PE 缓冲液（Qiagen）。14 000*g* 离心 1min。去除流过液，再 13 000r/min 离心 1min。将层析柱置于新的 1.5ml EP 小离心管。

（6）在层析柱膜的中心部位加入 35μl 的 USP 超纯水。室温孵育 2min（见注 6）。

（7）14 000*g* 离心 1min 洗提出 DNA。混合两个管中的洗提液。得到的 DNA 可以

立即进行大肠杆菌电转化实验，也可冻存以备后用。

(8) 同时与 dU-ssDNA 模板电泳 1.0μl 的上述洗提出的反应产物。用含有溴化乙锭（EB）的 TAE 琼脂糖凝胶电泳来检测 DNA（图 12.2 及注 7）。

一个成功的反应能使 dU-ssDNA 完全转化成低电泳迁移率的 dsDNA。通常，至少能看到两条产物带并且不存在 dU-ssDNA（图 12.2）。其中电泳迁移率较高的带就是所需的产物——正确延伸和连接的 CCC-dsDNA，可以用于有效转化大肠杆菌并且提供较高突变效率（约 80%）。而电迁移率较低的带是由 T7 DNA 聚合酶星号活性所致的单链异常产物[14]，其突变率较低（约 20%），转化效率也只有 CCC-dsDNA 的 1/30。如果 ssDNA 模板的重要部分转化为 CCC-dsDNA，那么就能得到一个高突变、高多样性的库。有时能得到电迁移率介于前述二者之间的第 3 条带，这条带被正确延伸但是没有连接的 dsDNA（图 12.2）。产生这一条带的原因要么是 T4 DNA 连接酶活性不够，要么就是寡核苷酸磷酸化不完全。

12.3.2.3 大肠杆菌电穿孔和噬菌体扩增

要完成库构建，异源 CCC-dsDNA 还必须引入含有 F′附加体的大肠杆菌宿主中。M13 噬菌体能侵染这种宿主并在其中增殖。噬菌体展示库的多样性还受限于 DNA 引入大肠杆菌的方法，其中高压电转化法的效率最高。

我们构建了一个大肠杆菌菌株——SS320，这是高效电转化和噬菌体增殖的理想菌株[2]。用标准的噬菌体杂交方案[15]，我们将 F′附加体从 *E. coli* XL-1 Blue（Stratagene）转入 *E. coli* MC1061（Bio-Rad）。*E. coli* MC1061 的染色体标记有链霉素抗性，而 *E. coli* XL-1 Blue 的附加体含有四环素抗性，这样杂交产生的子代菌株可以通过链霉素和四环素双抗进行选择。*E. coli* SS320 保留了 *E. coli* MC1061 的高电转化效率，同时 F′附加体的存在使得 M13 噬菌体感染宿主成为可能。

(1) 将上述纯化好的 DNA（来自 3），约 20μg 置于最小体积）及 0.2cm 间隙的电转槽置于冰上冷却。在冰上融化 350μl 分装好的电转感受态 *E. coli* SS320 细胞。将感受态细胞加入 DNA 中且用移液枪吸放数次混合（避免产生气泡）。

(2) 转移混合物于电转槽中进行电穿孔转化。电穿孔转化操作时，应遵循仪器厂家的指导手册，建议使用 BTX ECM-600 电转化仪时应设置：场强 2.5kV，电阻 129Ω，电容 50μF。也可以使用 BioRad 的 Gene Pulser 电转化仪及如下设置：场强 2.5kV，电阻 200Ω，电容 25μF。

(3) 立即在电穿孔转化后的电转槽中加入 1ml SOC 培养基以营救这些细胞并且转移培养基到一个 250ml 的锥形瓶中。用 1ml SOC 培养基冲洗电转槽两次。再向瓶中加入 SOC 培养基到终体积为 25ml，在 200r/min 摇床 37℃孵育 20min。

(4) 要确定所建库的多样性，可以在 LB/carb 培养皿中系列稀释涂板来选择适当的噬菌粒（如带有β-内酰胺酶基因的噬菌粒 pS1607）。

(5) 加入 M13KO7 helper 噬菌体（4×10^{10} pfu/ml），在 200r/min 摇床 37℃孵育 10min。

(6) 转移培养液于含有 500ml 2YT 培养基的 2L 锥形瓶中，加入合适的抗生素进行噬菌粒选择（如 2YT/carb 培养基）。

(7) 在 200r/min 摇床 37℃孵育 1h 且加入 25μg/ml 的卡那霉素。然后在 200r/min

摇动下 37℃孵育过夜。

(8) 将上述培养液在 4℃及 16 000*g* 下（10 000r/min 在 Sorvall GSA 转子中）离心 10min。转移上清液到含有 1/5 体积 PEG/NaCl 的新试管中以沉淀噬菌体。室温孵育 5min。

(9) 在 Sorvall GSA 转子中 4℃及 16 000*g* 下离心 10min。去除上清液。再简短离心一下，用移液管吸走上清液残余。重新悬浮噬菌体小团于 1/20 体积的 PBS 中。

(10) 在 4℃及 27 000*g* 下（在 Sorvall SS-34 转子中，15 000r/min）离心 5min 以去除不溶性杂质。转移上清液于一个干净试管。

(11) 利用分光光度计估计噬菌体浓度（268nm 的光密度 [OD_{268}] =1.0 时溶液中噬菌体浓度约为 5×10^{12} 噬菌体/ml；见注 8）。

12.3.2.4　制备 *E. coli SS320* 电穿孔法感受态

下述方案能产生大约 12ml 高浓度 *E. coli SS320* 电穿孔法感受态（约 3×10^{11} cfu/ml）。细胞可以长期储存于−70℃。

(1) 在 1ml 的 2YT/tet 培养基中接种一个生长于新鲜 LB/tet 培养皿的 *E. coli SS320* 菌株单克隆。在 200r/min 摇床 37℃孵育 6～8h。

(2) 转移培养液于有 500ml 的 2YT/tet 培养基的 2L 锥形瓶中，在 200r/min 摇动下 37℃孵育过夜。

(3) 用 5ml 上述过夜培养液接种 6 个 2L 锥形瓶，每瓶含有 900ml 的 Superbroth/tet 培养基（见 12.2.1.4）。在 200r/min 摇动下 37℃孵育到 OD_{550} 约为 0.8。

(4) 在冰上冷却 3 个上述锥形瓶 5min，不时摇动。下列步骤 (5) ～ (12) 应该在冷室中及冰上进行，所用一切溶液及仪器应该预冷。

(5) 在 Sorvall GS-3 转子中 4℃及 5000*g*（5500r/min）下离心 10min。去除上清液然后从其他 3 个瓶中加入余下的培养液（所有液体需要预冷）。重复上述离心及去除上清液步骤。

(6) 在离心瓶中加入 1.0mmol/L HEPES，pH7.4，同时加入灭菌的磁铁搅拌棒来帮助重悬离心沉淀的细胞。摇动使沉淀脱离管壁，并且在适中的速度下磁力搅拌以完全重悬细胞沉淀。

(7) 在 Sorvall GS-3 转子中 4℃及 5000*g*（5500r/min）下离心 10min。去除上清液，小心管中的磁铁搅拌棒。从转子中取出离心管时要小心保持离心管的位置以避免干扰管底的细胞沉淀。

(8) 在离心瓶中加入 1.0mmol/L HEPES，pH7.4。重悬细胞沉淀，重复步骤 (6) 和 (7) 中的重悬及离心过程。去除上清液。

(9) 重悬每管细胞沉淀于 150ml 的 10%超纯甘油中。不要混合各离心管。

(10) 在 Sorvall GS-3 转子中 4℃及 5000*g*（5500r/min）下离心 15min。去除上清液并且取出磁铁搅拌棒。用一个移液管小心吸出残余的上清液。

(11) 在一个离心管中加入 3.0ml 的 10%超纯甘油，用（移液管）小心抽吸重悬细胞沉淀。转移悬浮好的细胞到另一个离心管，重复上述过程直到所有细胞沉淀都得到很好地悬浮。

(12) 在液氮中快速冰冻分装为 350μl 的感受态细胞并且储存于−70℃。

12.3.3　选择和分析能提高融合蛋白展示水平的P8变异体

12.3.3.1　从hGH-P8库中选择噬菌体

将hGHbp包被于96孔Maxisorp免疫板上作为诱饵，上述hGH-P8库中的噬菌体通过多轮结合实验进行筛选[16]。在M13KO7辅助噬菌体的帮助下，噬菌体在*E. coli* XL-1 Blue中进行增殖以备后续筛选之用。

1）用靶蛋白包被96孔Maxisorp板

（1）用100μl 5μg/ml的靶蛋白（如hGHbp）溶液包被96孔Maxisorp板中的8孔，4℃孵育过夜。直接倾倒96孔板于水池，以去除孔中溶液。

（2）在每孔中加入200μl 0.2%溶于PBS的BSA溶液以阻止其他蛋白质对Maxisorp板的非特异性结合，室温下摇动1h。

2）噬菌体库的选择

（1）在上述每孔中加入一定量的置于PBS-T-BSA缓冲液中的噬菌体库（约10^{12} phage/ml）。在室温下摇动2h。然后用PBS-T缓冲液洗96孔Maxisorp板8次。结合筛选的严厉性（stringency）可以在以后的筛选中通过增加洗涤的次数来调控。

（2）在每孔中加入100μl的100mmol/L HCl以洗脱结合的噬菌体。在室温下强力摇动5min。

（3）将所有洗脱液收集到一起，加入1/5体积的1.0mol/L Tris-碱中和。

3）增殖噬菌体以备后用

（1）将上述洗脱出的噬菌体混合液加入到10倍体积的XL-1 Blue细胞中（OD_{550}＝0.5～1.0）。

（2）37℃下200r/min摇动孵育20min，取出10μl留备测量滴度，参考12.3.2.3节中步骤（4）。

（3）加入M13KO7 helper噬菌体，在200r/min摇床37℃孵育45min。

（4）转移培养液于100ml的2YT/carb/kan培养基中，在200r/min，37℃下摇动过夜。

（5）利用PEG/NaCl沉淀法分离噬菌体，参照12.3.2.3节，步骤（8）～（10）。

（6）重复上述筛选过程5次，每轮只用一半的洗脱噬菌体。

12.3.3.2　噬菌体酶联免疫法测定hGH展示水平

在hGH展示选择后（见12.3.3.1），通过酶联免疫法[2,11]可以测定单个克隆的hGH相对于P8的展示水平。顺次稀释的hGH-P8噬菌体溶液在含固定化hGHbp诱饵的板上孵育。洗去未结合的噬菌体，结合的噬菌体经过辣根过氧化物酶偶联的M13抗体反应后可以通过光谱学进行检测。从噬菌体浓度对450nm吸收值曲线上，通过比较噬菌体浓度和特定吸收值关系，可以估计出P8变异体较野生型P8对hGH展示水平提高的程度[11]。hGH展示水平高的克隆被送去测定其DNA序列，进而推断出融合有hGH的P8变异体的序列。

（1）从新鲜的LB/tet板上挑出一个含有特定噬菌粒的*E. coli* XL-1 Blue菌株的单克隆，放入1ml加有M13KO7 helper噬菌体（10^{10} pfu/ml）和50μg/ml羧苄青霉素（以保持

噬菌体）及 5μg/ml 四环素（以保持 F′附加体）的 2YT 培养基中。在 200r/min，37℃下摇动 2h，再加入 25μg/ml 的卡那霉素来选择共转染了 M13KO7 的克隆。再在 200r/min，37℃下摇动 6h，转移培养液于 30ml 2YT/carb/kan 培养基中。在 200r/min，37℃下摇动过夜。

（2）在 4℃及 27 000*g* 下（Sorvall SS-34 转子中，15 000r/min）离心 10min。转移上清液于一个含有 1/5 体积 PEG/NaCl 的干净试管中，室温孵育 5min。在 4℃及 12 000*g* 下（10 000r/min，Sorvall SS-34 转子中）离心 10min。去除上清液。在 2000*g*（4000r/min）简短离心一下，再用移液管吸走上清液残余。

（3）重悬噬菌体沉淀小团于 0.5ml 的 PBS-T-BSA 缓冲液中，将其转移到一个 1.5ml 的 EP 离心管中。在桌式离心机中用 14 000*g* 离心 5min，再转移上清液到一个新的 1.5ml EP 离心管中。

（4）利用分光光度计法估计噬菌体浓度 [$\varepsilon_{268}=1.2\times10^{8}$/(mol · L）/cm]。

（5）用 PBS-T-BSA 缓冲液准备 5 倍系列稀释的噬菌体储液。

（6）转移 100μl 的噬菌体溶液到有 hGHbp 包被且 BSA 封闭的 96 孔 Maxisorp 免疫板（见 12.3.3.1 中 1））。温和摇动孵育 1h。

（7）去除噬菌体溶液，用 PBS-T 缓冲液洗板 8 次。

（8）加入 100μl 的辣根过氧化物酶（horseradish peroxidase）/抗 M13 抗体络合物（用 PBS-T-BSA 缓冲液稀释 3000 倍，见 12.2.2.2）。温和摇动孵育 30min。

（9）用 PBS-T 缓冲液洗 8 次，再用 PBS 洗 2 次。

（10）用 100μl 的 TMB 底物溶液显影 96 孔免疫板。用 100μl 的 1.0mol/L H_3PO_4 溶液终止反应，然后在 96 孔板读板仪中读出 450nm 的光吸收值。

12.4　注

1. 通过突变引入终止密码子，方案可以缩减为 12.3.2.2 中的 2）和 12.3.2.3 中的 3）中的 1/10。反应产物可以直接用于转化 *E. coli*。

2. 这一方案可以通过过夜孵育大量产物以及用多次离心柱纯化 ssDNA 进行扩大化。

3. 除退火外，所有步骤都能相应进行扩大化。所述退火方案适用于 250μl 体积左右。对更大体积也许依然适用，但我们没有测试过。

4. 在随后步骤中，磷酸化的寡核苷酸应该立即使用。

5. 可以通过多个 250μl 体系退火反应来扩大化这一反应。

6. 洗脱时，加水后在 37℃孵育 5min 可以提高 DNA 的洗脱效率。

7. 环状 DNA 的电泳迁移率取决于盐浓度、pH 和 EB 的存在与否。观察图 12.2 中的相对迁移率，DNA 必须在直接加有 EB 的 TAE/琼脂糖胶上进行电泳，切忌将 EB 加到电泳缓冲液中。

8. 立即使用展示库或是加入 10%甘油储存在－80℃。冷冻能使有些展示蛋白变性，使之变得不稳定，不利于选择。一般而言，最好尽快使用展示库。

参考文献

1. Smith，G. P. and Petrenko，V. A.（1997）Phage display. *Chem. Rev.* 97，391-410.

2. Sidhu，S. S.，Lowman，H. B.，Cunningham，B. C.，and Wells，J. A.（2000）Phage display for selection of novel

binding peptides. *Methods Enzymol.* 328, 333-363.

3. Sidhu, S. S. (2000) Phage display in pharmaceutical biotechnology. *Curr. Opin. Biotechnol.* 11, 610-616.

4. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315-1317.

5. Bass, S., Greene, R., and Wells, J. A. (1990) Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins* 8, 309-314.

6. Malik, P., Terry, T. D., Gowda, L. R., et al. (1996) Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage. *J. Mol. Biol.* 260, 9-21.

7. Iannolo, G., Minenkova, O., Petruzzelli, R., and Cesareni, G. (1995) Modifying filamentous phage capsid: limits in the size of the major capsid protein. *J, Mol. Biol.* 248, 835-844.

8. Kretzschmar, T. and Geiser, M. (1995) Evaluation of antibodies fused to minor coat protein III and major coat protein VIII in bacteriophage M13. *Gene* 155, 61-65.

9. Clackson, T. and Wells, J. A. (1994) *In vitro* selection from protein and peptide libraries. *Trends Biotechnol.* 12, 173-184.

10. Weiss, G. A., Wells, J. A., and Sidhu, S. S. (2000) Mutational analysis of the major coat protein of M13 identifies residues that control protein display. *Protein Sci.* 9, 647-654.

11. Sidhu, S. S., Weiss, G. A., and Wells, J. A. (2000) High copy display of large proteins on phage for functional selections. *J, Mol. Biol.* 296, 487-495.

12. Roth, T. A., Weiss, G. A., Eigenbrot, C., and Sidhu, S. S. (2002) A minimized M13 coat protein defines the minimum requirements for assembly into the bacte-riophage particle. *J. Mol. Biol.* 322, 357-367.

13. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* 154, 367-382.

14. Lechner, R. L., Engler, M. J., and Richardson, C. C. (1983) Characterization of strand displacement synthesis catalyzed by bacteriophage T7 DNA polymerase. *J. Biol. Chem.* 258, 1174-1184.

15. Miller, J. H. (1972) *Experiments in Molecular Biology* 1st ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 190.

16. Fuh, G., Mulkerrin, M. G., Bass, S., et al. (1990) The human growth hormone receptor: secretion from *Escherichia coli* and disulfide bonding pattern of the extracellular binding domain. *J. Biol. Chem.* 265, 3111-3115.

17. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 27-28, 33-38.

18. Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* 15, 112-113, 132-134.

13　核糖体失活展示系统

Satoshi Fujita，Jing-Min Zhou，and Kazunari Taira

概要　在这里我们介绍一种在体外将基因型和表现型相联系的新策略，用来选择功能蛋白质。这个策略里，蓖麻蛋白的 A 链在翻译时被激活，因此不需要去掉终止密码子，就能形成核糖体、信使 mRNA 和翻译出蛋白质的稳定复合物。这个技术不需要转染、化学合成、连接，也不需要去掉终止密码子。因此，我们的新型核糖体失活展示系统能够为体外蛋白质的进化提供一个可靠的、简单的、完全的筛选系统，而不损失种群库，这样将随机和适度选择策略综合应用的方法，可以预测进化方向。
关键词　核糖体失活展示系统；体外选择；分子进化；蛋白质；核糖体展示；mRNA 展示；噬菌体展示；蓖麻蛋白 A 链；蛋白质-核糖体-mRNA 复合体

13.1　介　　绍

在过去的 10 年里，几种展示系统为多肽和蛋白质的选择及进化提供了强大而高效的技术。在这些技术中，将基因型和表现型相联系是最关键的决定因素。在这样的选择系统里，文库成员特异的序列信息（基因型）编码了被选择的蛋白质（表现型），它们是由引进这个系统的相应的 DNA/RNA 决定的。然后，编码被选择蛋白质的基因被再次扩增，以便进行下一轮的进化和筛选。这个策略被成功地发展起来：其一是细胞依赖型，包括展示在噬菌体的表面[1]、其他病毒的表面[2]、细菌的表面[3]、酵母的表面[4]；其二是无细胞型，如核糖体展示系统[5~7]以及信使 RNA（mRNA）展示系统[8,9]。

然而，现在发展起来的方法，都难免具有一些局限性和缺点。因为细胞依赖型展示系统[1~4]包括了必需的体内步骤、转化效率和蛋白质性质限制了序列文库的大小和多样性。例如，一些对细胞有害的蛋白质，或者在细胞中发挥重要调节功能的蛋白质就不能被选择。在核糖体展示系统中[5~7]，核糖体和翻译出的蛋白质、编码蛋白质的 mRNA 形成一个稳定的复合体，这是因为去掉密码子，添加 $Mg(OAc)_2$ 和反义-ssrA 多核苷酸[6,7]，可以使蛋白质从复合体中释放的速度减慢。但是，在一轮核糖体展示后脱离下来的 mRNA 量不是很高。用来替代的无细胞型 mRNA 展示程序[8,9]，需要精细地合成以及严格地纯化连接嘌呤霉素的寡聚核苷酸，这段核苷酸片段将被连接到 mRNA 文库序列的 3′端。任何不恰当的操作都将导致文库多样性的降低。

为了解决上述问题，保持序列文库的多样性，我们发展了一种新策略，得到蛋白质-核糖体-mRNA 三元复合体，它比用传统方法得到的复合体更稳定（图 13.1）。这种稳定性是通过在被翻译成蛋白质文库的序列文库的下游区域引进一种可以使真核细胞核糖体失活的毒素基因——蓖麻蛋白 A 链（RTA）来实现的（图 13.1）。蓖麻蛋白是由 A 链、B 链组成的植物毒素，它能使核糖体失活以抑制蛋白质的合成。B 链是蓖

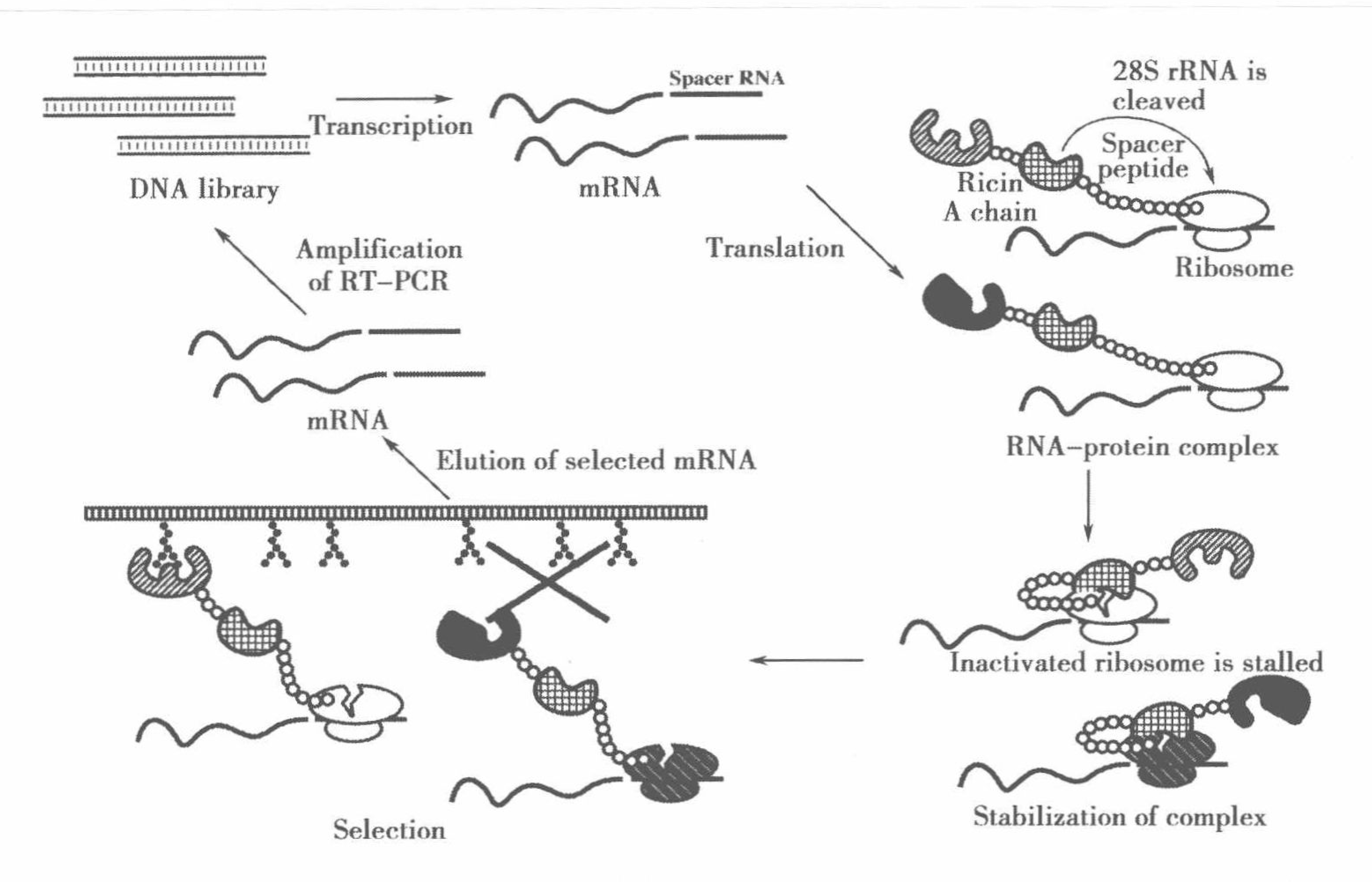


图 13.1 RIDS 体外筛选功能蛋白质的流程示意图。第一步，转录：RTA 的基因被插入到随机蛋白质文库（或者 cDNA 文库）的下游，在 T7 RNA 聚合酶的作用下产生 mRNA。第二步，翻译：在兔网织红细胞裂解系统中，mRNA 被翻译。第三步，失活：在翻译时，因为在顺式反应里，折叠的 RTA 使 rRNA 失活，核糖体的功能丧失，形成核糖体-mRNA-蛋白质复合体。第四步，选择：目标复合体结合到相应的亲和基质上。不需要的复合物则被洗脱下来。第五步，洗脱：用含有乙二胺四乙酸（EDTA）的缓冲液洗脱，使特异复合物从基质上解离下来，空出的 mRNA 也分离下来。第六步，扩增：用 RT-PCR 方法扩增洗脱下来的 RNA，因此产生的 cDNA 进入下一个循环或者进行克隆测序分析。

麻蛋白通过内化作用进入细胞时必需的。真核核糖体大亚基的 23S～28S 核糖体 RNA（tRNA）上有一个 GAGA 四碱基环，蓖麻蛋白的活性部分 RTA 催化水解邻近其中一个普遍保守的腺嘌呤的 *N*-糖苷键[10,11]。这个单点脱嘌呤改变了与延伸因子的结合位点，抑制了蛋白质合成，使翻译复合体上的核糖体还没有释放 mRNA 或翻译蛋白就失活[12～14]。由于翻译在到达终止密码子之前就停止了，核糖体就不用招募释放因子。因此，mRNA 和初生蛋白质仍旧非常稳定地结合在核糖体上，不只是数小时，而且是数天。特别的是，RTA 每分钟可以使约 1500 个核糖体失活[15]，因此，在标准的翻译条件下能够快速形成极其稳定的蛋白质-核糖体-mRNA 复合体。所以，核糖体失活展示系统（RIDS）的关键之处就在于，mRNA 上的核糖体通过与原先的核糖体展示系统完全不同的机制而失活。利用这种方法，我们能够在标准的翻译条件下筛选功能蛋白，而不需要去除终止密码子。

为了评估 RIDS 潜在的应用价值，我们选择链霉亲和素和谷胱甘肽转移酶（GST）作为目标蛋白，用分别含有生物素或谷胱甘肽做配体的基质分离各自的蛋白质和 mRNA。我们发现，利用 RIDS 达到目的是可能的，即在不降低序列文库多样性，不需要化学合成的情况下，筛选出功能蛋白（链霉亲和素和 GST）。如果向 RIDS 中引进互补的 DNA（cDNA）文库或者随机 DNA 文库替代模式蛋白（链霉

亲和素或 GST)，那么我们就可以选择出功能蛋白。

13.2 材　　料

(1) pET30a (Novagen, Darmstadt, Germany)。

(2) cDNA 文库或随机 DNA 文库。

(3) 链霉亲和素和 GST 基因。

(4) RTA 基因。

(5) 来源于 pCANTAB 5E (Amersham Biosciences, Piscataway, NJ) 的 fd 噬菌体基因Ⅲ。

(6) 多核苷酸引物和连接子。

(7) T7 Ampliscribe™ 试剂盒 (Epicentre Technologies, Madison, WI)。

(8) 帽子结构的类似物 (New England Biolabs, England)。

(9) DNase I (Epicentre Technologies, Madison, WI)。

(10) RNeasy™ mini 试剂盒 (QIAGEN, Hilden, Germany)。

(11) Flexi® 兔网织红血球裂解系统 (Promega, Madison, WI)。

(12) Recombinant RNasin® 核糖核酸酶抑制剂 (Promega)。

(13) 结合 (或漂洗) 缓冲液：140mmol/L NaCl、2.7mmol/L KCl、10.1mmol/L Na_2HPO_4、1.8mmol/L KH_2PO_4、0.04% Tween-20、2% Block Ace™ (Dainippon Pharmaceutical Co.; Japan)、5mmol/L $MgCl_2$，pH7.3。可以在−20℃储存一个月。

(14) 选择用的配体珠子。

(15) 生物素琼脂糖 (Sigma, St. Louis, MO)、谷胱甘肽凝胶 4B (Amersham Biosciences)。

(16) 洗脱缓冲液：1mol/L NaCl、50mmol/L EDTA。

(17) ReverTra Dash® (Toyobo, Japan)。

(18) KOD Dash® (Toyobo)。

13.3 方　　法

这里介绍的方法概括了表达载体的构建以及筛选功能蛋白的 RIDS 循环系统的构建。

13.3.1 RIDS 表达载体的构建

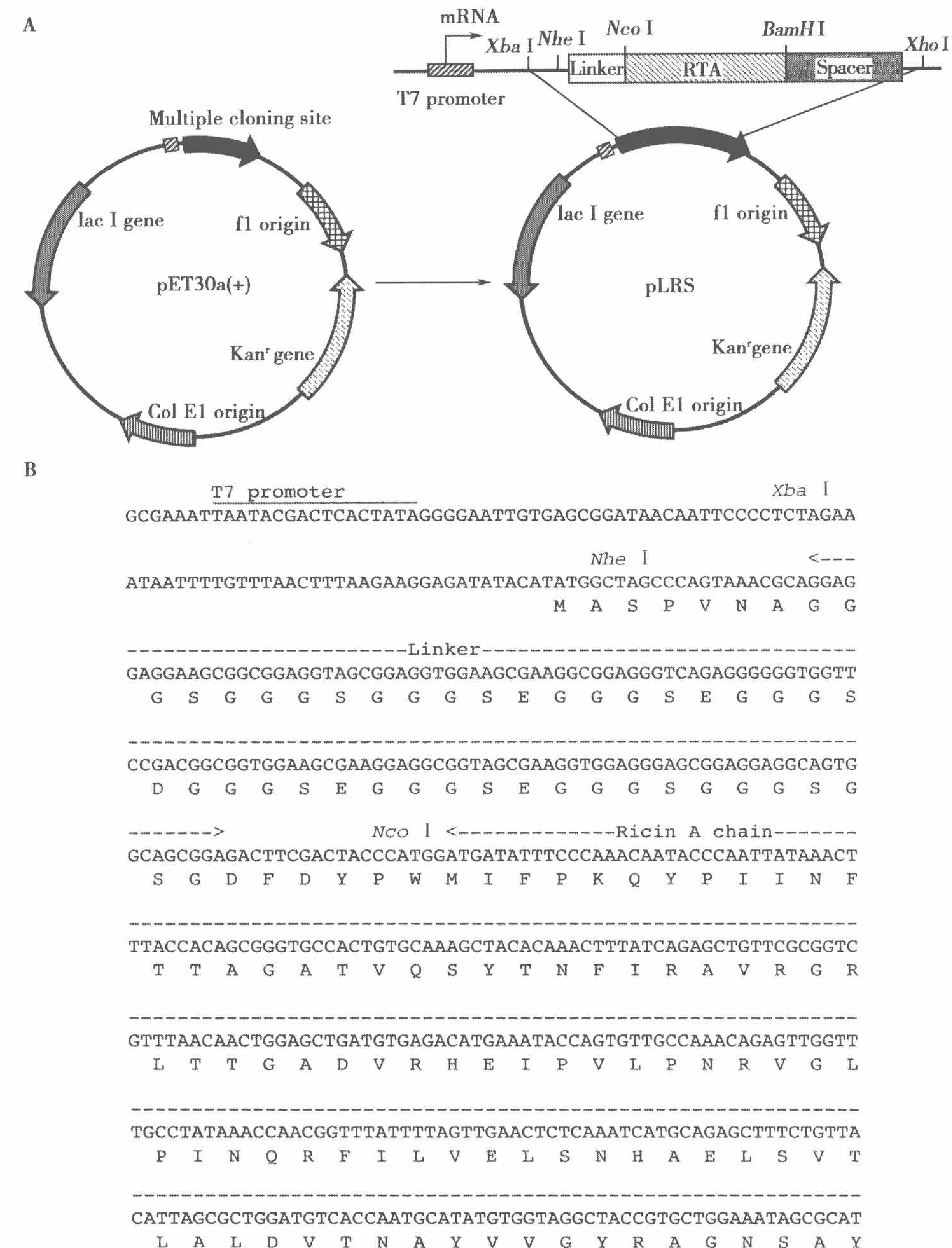
首先，需要准备含编码 T7 启动子、蛋白质文库、连接子、RTA 和间隔区的 DNA 来构建 RIDS (图 13.2A)。13.3.1.1～13.3.3.3 部分详述了这个过程。尽管理论上用聚合酶链反应 (PCR) 制备双链 DNA 是可行的，但是我们推荐用质粒来构建此 DNA，因为用 PCR 来连接不同部分 (包括 T7 启动子、蛋白质文库、连接子、RTA 和间隔区)，很难保证不会引入非特异性扩增。在这个章节中，我们描述了构建 RIDS 载体的过程。由于采用标准的重组 DNA 方法来构建载体，因为篇幅有限，DNA 的操作过程在此就不赘述了。

13.3.1.1 构建 pLRS 表达载体

首先，我们需要准备通用载体 pLRS，它像构建 RIDS 的 DNA 一样，能编码 T7 启动子、蛋白质文库、连接子、RTA 和间隔区 (图 13.2A)。这个载体是基于 pET30a 载

体（Novagen）的表达系统。我们通过将 cDNA 文库或者随机文库插入到 pLRS 载体的 *Xba*I/ *Nhe*I 位点可以很容易地构建 RIDS 蛋白质文库。

用 DNA 合成仪合成连接子片段，并通过 T7 启动子下游的 *Xba*I/ *Xho*I 位点插入到 pET30a 载体中。连接子编码了 *Nhe*I、噬菌体 M13 基因Ⅲ的甘氨酸/丝氨酸富集序列、*Nco*I、*Bam*HI、*Pst*I 和 *Xho*I（顺序从上游开始）。编码 RTA 的 DNA 是通过 PCR 从pUTA质粒中（获赠于得克萨斯大学的 J. Robertus 教授）得到的，并且将其通过 *Nco*I/ *Bam*HI 位点插入到甘氨酸/丝氨酸富集序列的下游。编码间隔区的 DNA 来源于噬菌体 M13 基因Ⅲ，并且通过 *Bam*HI/ *Pst*I 位点插入到 RTA 的下游。DNA 的序列（不包括插入的 RIDS 蛋白质文库序列）如图 13.2B 所示。



```
-----------------------------------------------------------
ATTTCTTTCATCCTGACAATCAGGAAGATGCAGAAGCAATCACTTATCTTTTCACTGATG
 F  F  H  P  D  N  Q  E  D  A  E  A  I  T  Y  L  F  T  D  V

-----------------------------------------------------------
TTCAAAATCGATATACATTCGCCTTTGGTGGTAATTATGATAGACTTGAACAACTTGCTG
 Q  N  R  Y  T  F  A  F  G  G  N  Y  D  R  L  E  Q  L  A  G

-----------------------------------------------------------
GTAATCTGAGAGAAAATATCGAGTTGGGAAATGGTCCACTAGAGGAGGCTATCTCAGCGC
 N  L  R  E  N  I  E  L  G  N  G  P  L  E  E  A  I  S  A  L
-----------------------------------------------------------
TTTATTATTACAGTACTGGTGGCACTCAGCTTCCAACTCTGGCTCGTTCCTTTATAATTT
 Y  Y  Y  S  T  G  G  T  Q  L  P  T  L  A  R  S  F  I  I  C

-----------------------------------------------------------
GCATCCAAATGATTTCAGAAGCAGCAAGATTCCAATATATTGAGGGAGAAATGCGCACGA
 I  Q  M  I  S  E  A  A  R  F  Q  Y  I  E  G  E  M  R  T  R

-----------------------------------------------------------
GAATTAGGTACAACCGGAGATCTGCACCAGATCCTAGCGTAATTACACTTGAGAATAGTT
 I  R  Y  N  R  R  S  A  P  D  P  S  V  I  T  L  E  N  S  W

-----------------------------------------------------------
GGGGGAGACTTTCCACTGCAATTCAAGAGTCTAACCAAGGAGCCTTTGCTAGTCCAATTC
 G  R  L  S  T  A  I  Q  E  S  N  Q  G  A  F  A  S  P  I  Q

-----------------------------------------------------------
AACTGCAAAGACGTAATGGTTCCAAATTCAGTGTGTACGATGTGAGTATATTAATCCCTA
 L  Q  R  R  N  G  S  K  F  S  V  Y  D  V  S  I  L  I  P  I

--------------------Ricin A chain--------------->
TCATAGCTCTCATGGTGTATAGATGCGCGCCTCCACCGAGCTCACAGTTTGGGATATCGT
 I  A  L  M  V  Y  R  C  A  P  P  P  S  S  Q  F  G  I  S  S

                                          <--------------------
                                         BamH I
CGACCGGAGGAGGAGGTGGCGGCGGTGGCGCTGCATCGGATCCATTCGTTTGTGAATATC
 T  G  G  G  G  G  G  G  A  A  S  D  P  F  V  C  E  Y  Q

-----Spacer (gene III)--------------------------------------
AAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGTG
 G  Q  S  S  D  L  P  Q  P  P  V  N  A  G  G  G  S  G  G  G

-----------------------------------------------------------
GTTCTGGTGGCGGCTCTGAGGGTGGCGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCT
 S  G  G  G  S  E  G  G  G  S  E  G  G  G  S  E  G  G  G  S

-----------------------------------------------------------
CTGAGGGTGGCGGTTCCGGTGGCGGCTCCGGTTCCGGTGATTTTGATTATGAAAAAATGG
 E  G  G  G  S  G  G  G  S  G  S  G  D  F  D  Y  E  K  M  A

-----------------------------------------------------------
CAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGACG
 N  A  N  K  G  A  M  T  E  N  A  D  E  N  A  L  Q  S  D  A

-----------------------------------------------------------
CTAAAGGCAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCTATCGATGGTTTCATTG
 K  G  K  L  D  S  V  A  T  D  Y  G  A  A  I  D  G  F  I  G

----------------Spacer (gene III)------->                Xho
GTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTCTGCAGGTTAAGAATTCTC
 D  V  S  G  L  A  N  G  N  G  A  T  G  L  Q  V  K  N  S  R

 I
GAGCACCACCACCACCACCAC
```

图 13.2　A. pET30a（＋）和 pLRS 质粒示意图。我们可以很容易地将 cDNA 文库或者随机文库插入到 pLRS 载体的 *Xba*I/ *Nhe*I 位点来构建 RIDS 蛋白质文库。B. RIDS 使用的 pLRS 的 DNA 序列。T7 启动子、蛋白质文库、连接子、RTA 和间隔区的序列如图所示。

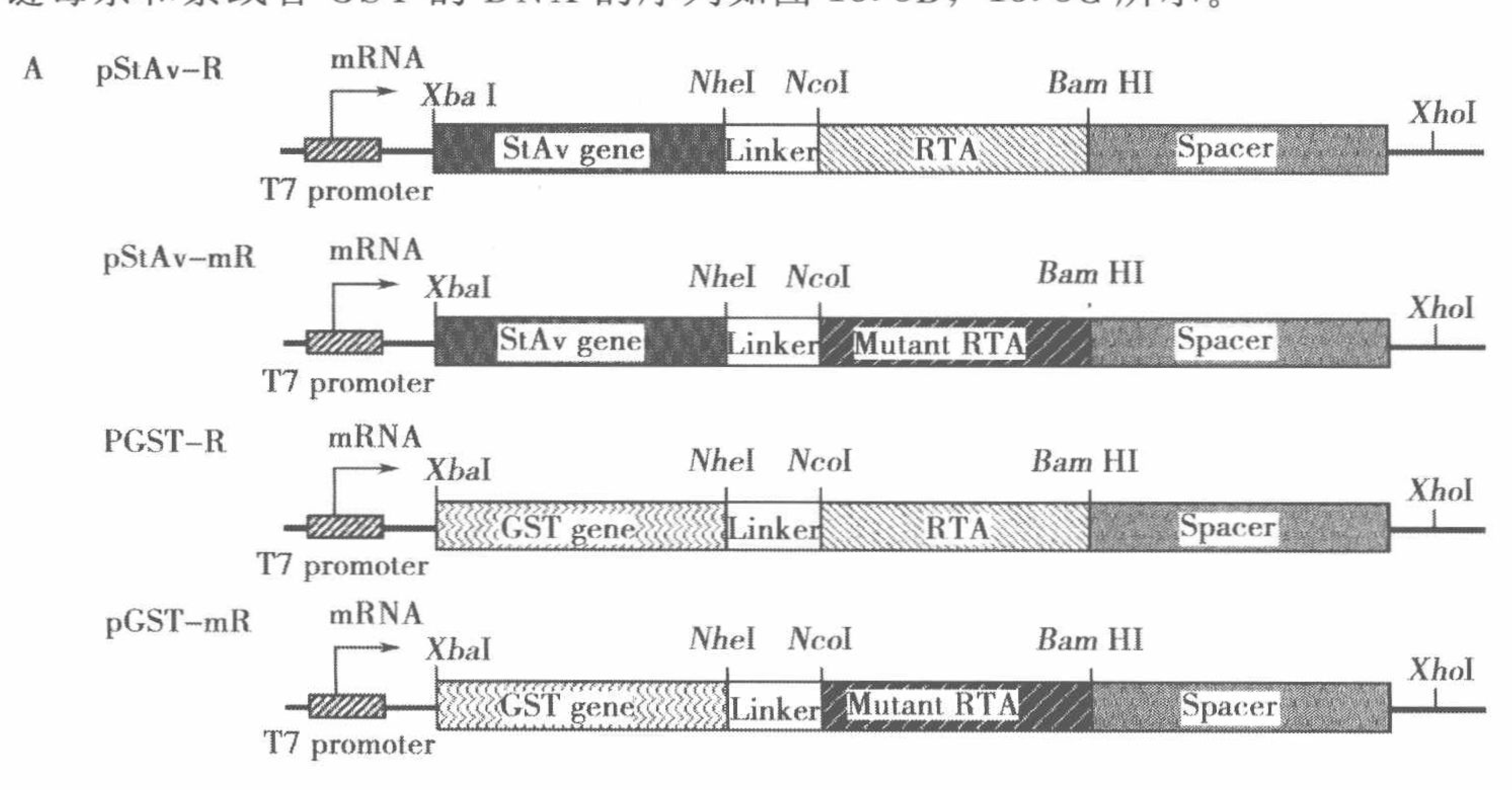
T7 启动子和蛋白质文库第一个 ATG 之间的 71-mer 序列来自于 pET30a 载体。RTA 上游灵活的连接子对于初生蛋白折叠成正确的三维结构，去除蛋白质文库和下游 RTA 之间的空间阻隔是必需的。之前的研究表明，被选择的蛋白质/mRNA 量强烈依赖于这个连接子的长度、组成和序列[6,16,17]。在目前的研究中，我们用 44 个含有甘氨酸/丝氨酸富集序列的氨基酸作为连接子。

去除 RTA 的空间阻隔是非常重要的。这是因为，在 RIDS 中通过引进 RTA 基因来达到稳定 mRNA-核糖体-蛋白质复合体的目的。除了这个连接子之外，可读框（ORF）3′端的间隔区也是非常重要的。这个间隔区作为锚来连接核糖体，必须有合适的长度才能占据核糖体的长通道[18,19]，并且允许初生 RTA 正确折叠而没有任何空间障碍[20,21]。已经有报道，C 端至少需要 20～30 个氨基酸的间隔区来保持展示在核糖体上的酶的活性[22,23]。因此，核糖体展示系统上分离出的 mRNA 量会受到间隔子的长度，及其 3′端二级结构的影响[6,18,24]。我们在 ORF 的 3′端分别引进一个长的和一个短的间隔区序列，比较它们在翻译和选择方面的影响。我们发现，在我们的系统里，长的间隔子（由噬菌体 M13 的全长基因Ⅲ编码的 404 个氨基酸）对于翻译和选择是不合适的。所以，在我们所有的实验中，我们用 121 个氨基酸的片段作为间隔子。

13.3.1.2 pStAv-R 和 pGST-R 载体的构建

为了校正提出的方法和 RIDS 的潜在应用性，我们在 DNA 文库中引进编码链霉亲和素或 *GST* 基因作为模式研究（图 13.3 A）。这些蛋白质经常被融合进新发现的蛋白质或者与之相互作用的分子上，成功地研究新发现的融合蛋白的功能性质[25～28]。链霉亲和素和 GST 分别结合到生物素和谷胱甘肽上，这种结合具有高亲和性和特异性。因此，分离和确认蛋白质-核糖体-mRNA 复合体可能是很容易的。

从 pSTA（冈山大学的 M. Sisido 教授所赠）和 pGEX-4T-3（Amersham Biosicences）质粒上分别切除编码链霉亲和素和 GST 的 DNA。用 PCR 方法将这些片段进行扩增，并分别连接到 pLRS 质粒的 *Xba*I/*Nhe*I 位点中，构建成 pStAv-R 和 pGST-R 质粒。编码链霉亲和素或者 GST 的 DNA 的序列如图 13.3B，13.3C 所示。



B

```
       T7 promoter                                      Xba I
GCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAA

ATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCGAGCATGACTGGTGGACAGC
                                       M  A  S  M  T  G  G  Q  Q

                             <--------Streptavidin gene--------------
AAATGGGTACCGAATTCCATATGGACCCGTCCAAGGACTCCAAAGCTCAGGTTTCTGCAG
 M  G  T  E  F  H  M  D  P  S  K  D  S  K  A  Q  V  S  A  A

------------------------------------------------------------
CCGAAGCTGGTATCACTGGCACCTGGTATAACCAACTGGGGTCGACTTTCATTGTGACCG
 E  A  G  I  T  G  T  W  Y  N  Q  L  G  S  T  F  I  V  T  A

------------------------------------------------------------
CTGGTGCGGACGGAGCTCTGACTGGCACCTACGAATCTGCGGTTGGTAACGCAGAATCCC
 G  A  D  G  A  L  T  G  T  Y  E  S  A  V  G  N  A  E  S  R

------------------------------------------------------------
GCTACGTACTGACTGGCCGTTATGACTCTGCACCTGCCACCGATGGCTCTGGTACCGCTC
 Y  V  L  T  G  R  Y  D  S  A  P  A  T  D  G  S  G  T  A  L

------------------------------------------------------------
TGGGCTGGACTGTGGCTTGGAAAAACAACTATCGTAATGCGCACAGCGCCACTACGTGGT
 G  W  T  V  A  W  K  N  N  Y  R  N  A  H  S  A  T  T  W  S

------------------------------------------------------------
CTGGCCAATACGTTGGCGGTGCTGAGGCTCGTATCAACACTCAGTGGCTGTTAACATCCG
 G  Q  Y  V  G  G  A  E  A  R  I  N  T  Q  W  L  L  T  S  G

------------------------------------------------------------
GCACTACCGAAGCGAATGCATGGAAATCGACACTAGTAGGTCATGACACCTTTACCAAAG
 T  T  E  A  N  A  W  K  S  T  L  V  G  H  D  T  F  T  K  V

---------------------------------------------Streptavidin gene-
TTAAGCCTTCTGCTGCgAGCATTGATGCTGCCAAGAAAGCAGGCGTAAACAACGGTAACC
 K  P  S  A  A  S  I  D  A  A  K  K  A  G  V  N  N  G  N  P

-------------------->Nhe I
CTCTTGACGCTGTTCAGCAAGCTAGC
 L  D  A  V  Q  Q
```

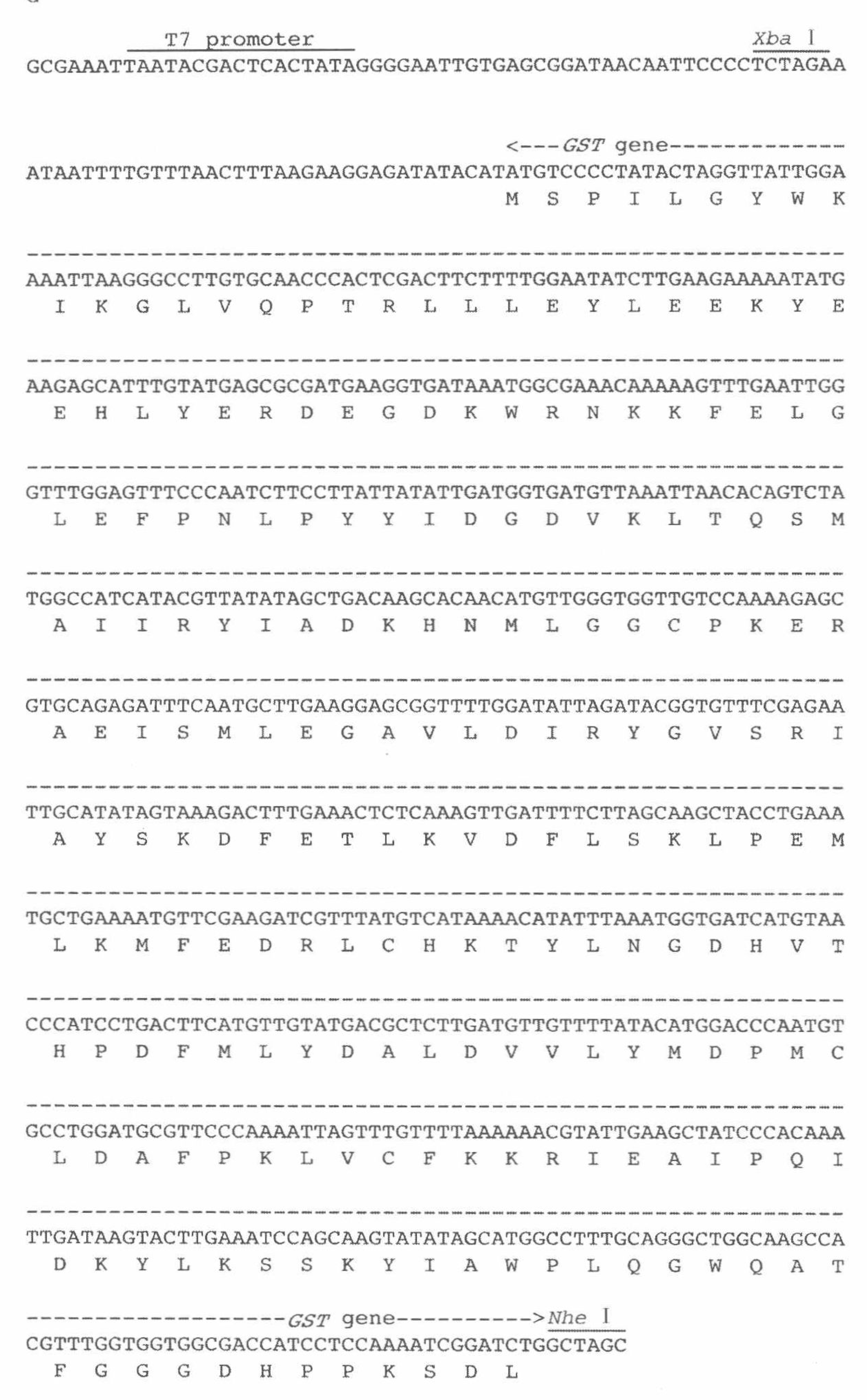


图 13.3 A. 作为模式研究的 RIDS 的 DNA 组成示意图。链霉亲和素和 *GST* 基因插入到 pLRS 质粒的 *Xba*I/ *Nhe*I 位点中，构建成 pStAv-R 和 pGST-R 质粒。为了确认 RTA 有利于稳定核糖体复合体，在对照实验中，我们通过点特异突变的方法得到 RTA 的突变失活基因，构建了 pStAv-mR 和 pGST-mR 质粒。B. RIDS 使用的 pStAv-R DNA 序列。C. RIDS 使用的 pGST-R DNA 序列。

13.3.1.3 pStAv-mR 和 pGST-mR 质粒的构建

为了确认 RTA 有利于稳定核糖体复合体，在对照实验中，我们通过点特异突变的方法改变了 RTA 功能氨基酸，得到 RTA 的突变失活基因。步骤如下：第 177 位的谷

氨酸突变成谷氨酰胺；第 180 位的精氨酸突变成组氨酸；第 208 位的谷氨酸突变成天冬氨酸。这 3 个氨基酸的突变完全抑制了 RTA 的活性[29,30]。为了构建 pStAv-mR 和 pGST-mR 质粒，每个质粒（pStAv-R 或 pGST-R）都采取标准的方法用 RTA 突变体基因替代 *RTA* 基因[31]，引物如下：5′-TTG CAT CCA AAT GAT TTC ACA AGC AGC ACA CTT CCA ATATAT TAG GGA GAA ATG-3′（下画线为 E177A 和 R180H 突变），以及 5′-GAT CCT AGC GTA ATT ACA CTT GAC GAT CCT AGC GTA ATT ACA CTT GA-3′（下画线为 E208D 突变）。

13.3.2　选择功能性蛋白的 RIDS 循环

采取标准的重组 DNA 方法，在含有随机文库或 cDNA 文库（在我们的模式研究中，是 pStAv-R、pGST-R、pStAv-mR 和 pGST-mR）的 ORF 中，用 *Xho*I 酶切 ORF 的 3′端产生线性 DNA。这个酶切反应是终止转录必需的。

13.3.2.1　DNA 文库到 mRNA 的转录

在 T7 RNA 聚合酶的作用下，DNA 被转录，含有一个帽状结构类似物。在我们的模式研究中，在供应商（Epicentre Technologies Co.）的推荐下，用 T7 Ampliscribe™ 试剂盒转录 DNA。

（1）根据供应商的推荐，向 20μl 转录混合物中添加 1μg 线性 DNA。转录混合物含有 3mmol/L 帽状结构类似物、7.5mmol/L rATP、rCTP、rUTP；0.75mmol/L rGTP；10mmol/L 反应缓冲液以及 AmpliScribe T7 酶溶液。

（2）在 37℃静置反应混合液 2～4h，产生各种 mRNA（我们将这些 mRNA 分别命名为：StAv-R、pGST-R、pStAv-mR 和 GST-mR）。

（3）在转录反应后，用 DNase I 完全消化模板 DNA，因为残余的模板 DNA 会影响后面的选择步骤。

（4）根据供应商（Qiagen）的推荐，用 RNeasy™ mini 试剂盒纯化 mRNA。

13.3.2.2　核糖体-mRNA-蛋白质复合体的亲和选择和 mRNA 的分离

制备 RNA 编码文库（在我们的模式实验中，是 StAv-R、pGST-R、pStAv-mR 和 GST-mR）和 Flexi 兔网织红细胞裂解系统（Promega；见注 1）。

（1）向 40μl 翻译混合物中添加 2μg mRNA。翻译混合物含有 33μl Flexi 兔网织红细胞裂解液、40mmol/L KCl、40μmol/L 总氨基酸混合液、40U Recombinant RNasin 核糖核酸酶抑制剂（Promega），用蒸馏水定容体积到 50μl。不要添加 DTT 和 Mg^{2+} 离子（见注 1）。

（2）在 30℃静置 20min。在我们的研究中，3 种 mRNA 体系被翻译（每种体系用 2μg RNA，它们分别是：以 1∶1 比例混合的 mRNA，以及作为对照的编码链霉亲和素或者 GST（StAv-R 和 GST-R）的 mRNA；见注 2）。

（3）翻译之后，添加 1ml 适当的结合缓冲液［见 13.2（13）］和配基固定珠子。在我们的研究里，添加 10μl 生物素琼脂糖或者谷胱甘肽珠子的终止液（见注 3）。

（4）在 4℃静置 1h，温和地搅拌有利于结合反应。尽管静置在室温下也是可以的，但是我们推荐 4℃。因为在反义转录 PCR（RT-PCR）里有时会检测到非特异性的结合（见注 4）。

（5）在结合反应后，用 500*g* 的转速离心 1min 来沉淀配基固定珠子。

(6) 去掉含有未结合 mRNA 和蛋白质的上清液。

(7) 添加 200μl 漂洗液［见 13.2 (13)］，温和地混合 30s。

(8) 用 500*g* 的转速离心 1min 来沉淀配基固定珠子。

(9) 去掉含有未结合 mRNA 和蛋白质的上清液。

(10) 重复漂洗步骤 (7～9) 2 次或者 3 次。

(11) 添加 100μl 洗脱缓冲液［见 13.2 (16)］，室温用力地摇动 30min，从沉淀配基固定珠子上分离出结合的 mRNA。

(12) 根据供应商的推荐，用 RNeasy™ 试剂盒纯化洗脱下来的 mRNA，用真空泵浓缩纯化后的 mRNA。

13.3.2.3 反义转录

用 RT 进行反义转录反应。在我们的模式研究中，根据供应商的推荐，用 ReverTra Ace (Toyobo) 进行反义转录反应。

(1) 将纯化后的 mRNA、4μl 10μmol/L RT 引物、每种 10mmol/L dUTP 2μl、4μl 5×RT 缓冲液，用蒸馏水定容到 18μl，根据供应商的推荐，不加 RNase 抑制剂和 ReverTra Ace。RT 引物用来 (5′-GTG TAG CTT TGC ACA GTG GC-3′) 识别 RTA 序列的上游部分。

(2) 将混合物在 65℃下放置 5min 变性后，迅速放到冰上冷却。

(3) 向混合物中添加 1μl RNase 抑制剂和 1μl ReverTra Ace。

(4) 42℃静置 1h 进行反转录。然后 99℃静置 5min 将酶失活。

13.3.2.4 聚合酶链反应

用热稳定 DNA 聚合酶进行 PCR 反应。在我们的模式研究中，根据供应商的推荐，用 KOD Dash (Toyobo) 进行 PCR 反应。

(1) 将 20μl cDNA、10μl 10×PCR 缓冲液、每种 2mmol/L dUTP 10μl、1μl 10mmol/L RT (作为下游引物)，以及 2μl 10μmol/L 上游引物，用蒸馏水定容到 99μl，不添加 KOD Dash。上游引物 (5′-AAT TTT GTT TAA CTT GAA GGA G-3′) 用来识别 T7 启动子序列的下游部分。

(2) 将混合物在 98℃下放置 2min 变性后，迅速放到冰上冷却。

(3) 添加 0.5μl KOD Dash。

(4) 运行 PCR 程序 (98℃ 1min；然后 98℃ 10s，55℃ 2s，72℃ 30s，循环 15～30 次；最后 72℃ 10min)。98℃ 静置 1min。

13.4 注

1. 与标准的兔网织红血球裂解系统相比，Flexi 兔网织红血球裂解系统提供了更灵活的反应条件。在展示含有二硫键的蛋白质时要去掉还原剂 (包括 DTT)。根据核糖体展示方法优化 Mg^{2+} 和 K^+ 的浓度[32]。

2. 我们发现有时亲和选择 GST 时，若用谷胱甘肽凝胶，会检测到高背景。但是，若用抗 GST 抗体替代谷胱甘肽凝胶，则没有背景。这个结果表明可能因为空间阻隔的存在，

在核糖体上展示的几个 GST 不能二聚化，而仅有的 GST 二聚体能够识别谷胱甘肽。

3. 添加 5～50mmol/L $MgCl_2$ 能够稳定核糖体复合体。

4. 尽管我们能够在室温下从 GST 和链霉亲和素的混合物中成功选择出 GST 或者链霉亲和素，但是有时候能够检测到非特异性结合。似乎冷却翻译混合体系能够稳定核糖体复合体。惊奇的是，虽然存在终止子，但是向结合缓冲液中添加 $MgCl_2$、冷却翻译混合物，能使少量 StAav-mR mRNA 结合到生物素琼脂糖上。似乎在存在终止子的核糖体展示系统中形成了 mRNA-核糖体-蛋白质复合体。

致谢 感谢得克萨斯大学的 J. D. Robertus 教授赠送 pUTA 质粒；冈山大学的 T. Hosaka 博士赠送 pGSH 质粒。感谢 Taira 实验室 Loura Nelson 博士的建议和讨论。这项研究是由 Promotion of Basic Research Activities 生物科学创新基金，以及 Japan. S. F. 的经济贸易工业部的科学技术进步奖学金资助的。我们最近通过利用串联重复融合的 MS2 包被蛋白（MSp）二聚体与对应特异结合位点、C 变体（Cv）的 RNA 序列的相互作用，发展了一种似乎更为有效的选择系统。对应特异结合位点和 C 变体允许在室温条件下进行选择，增强了 mRNA-核糖体-蛋白质复合体的稳定性。我们称这个系统为拥有增强结合的先进核糖体展示系统（ARiSA）[33～35]。

参考文献

1. Winter, G., Griffiths, A. D., Hawkins, R. E., and Hoogenboom, H. R. (1994) Making antibodies by phage display technology. *Annu. Rev. Immunol.* 12, 433-455.
2. Kasahara, N., Dozy, A. M., and Kan, Y. M. (1994) Tissue-specific targeting of retroviral vectors through ligand-receptor interactions. *Science*, 266, 1373-1376.
3. Georgiou, G., Poetschke, H. L., Stathopoulos, C., and Francisco, J. A. (1993) Practical applications of engineering Gram-negative bacterial cell surfaces. *Trends Biotechnol.* 11, 6-10.
4. Kieke, M. C., Cho, B. K., Boder, E. T., Kranz, D. M., and Wittrup, K. D. (1997) Isolation of anti-T cell receptor scFv mutants by yeast surface display. *Protein Eng.* 10, 1303-1310.
5. Mattheakis, L. C., Bhatt, R. R., and Dower, W. J. (1994) An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. USA* 91, 9022-9026.
6. Hanes, J. and Plückthun, A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* 94, 4937-4942.
7. He, M. Y., Menges, M., Groves, M. A. T., et al. (1999) Selection of a human antiprogesterone antibody fragment from a transgenic mouse library by ARM ribosome display. *J, Immunol. Methods* 231, 105-117.
8. Roberts, R. W. and Szostak, J. W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA* 94, 12297-12302.
9. Nemoto, N., Miyamoto-Sato, E., Husimi, Y., and Yanagawa, H. (1997) In vitro virus: Bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*. *FEBS Lett.* 414, 405-408.
10. Endo, Y. and Tsurugi, K. (1987) RNA *N*-glycosidase activity of ricin a chain. Mechanism of action of the toxic lectin ricin on eukaryotic ribosomes. *J. Biol. Chem.* 262, 8128-8130.
11. Endo, Y., Mitsui, K., Motizuki, M., and Tsuruki, K. (1987) The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28S ribosomal RNA caused by the toxins. *J. Biol. Chem.* 262, 5908-5912.
12. Moazed, D., Robertson, J., and Noller, H. (1989) Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S rRNA. *Nature* 334, 362-364.
13. Kudlicki, W., Kitaoka, Y., Odom, O. W., Kramer, G., and Hardesty, B. (1995) Elongation and folding of

nascent ricin chains as peptidyl-tRNA on ribosomes: the effect of amino acid deletions on these processes. *J. Mol. Biol.* 252, 203-212.

14. Munishkin, A. and Wool, I. G. (1997) The ribosome-in-pieces: binding of elongation factor EF-G to oligoribonucleotides that mimic the sarcin/ricin and thiostrepton domains of 23S ribosomal RNA. *Proc. Natl. Acad. Sci. USA* 94, 12280-12284.
15. Eiklid, K., Olsnes, S., and Pihl, A. (1980) Entry of lethal doses of abrin, ricin, and modeccin into the cytosol of HeLa cells. *Exp. Cell Res.* 126, 321-326.
16. Mössner, E., Koch, H., and Plückthun, A. (2001) Fast selection of antibodies without antigen purification: adaptation of the protein fragment complementation assay to select antigen-antibody pairs. *J. Mol. Biol.* 308, 115-122.
17. Liu, R., Barrick, J. E., Szostak, J. W., and Roberts, R. W. (2000) Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. *Methods Enzymol.* 318, 268-293.
18. Malkin, L. I. and Rich, A. (1967) Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J. Mol. Biol.* 26, 329-346.
19. Smith, W. P., Tai, P. C., and Davis, B. D. (1978) Interaction of secreted nascent chains with surrounding membrane in bacillus subtilis. *Proc. Natl. Acad. Sci. USA* 75, 5922-5925.
20. Komar, A. A., Kommer, A., Krasheninnikov, I. A., and Spirin, A. S. (1997) Cotranslational folding of globin. *J. Biol. Chem.* 272, 10646-10651.
21. Fedorov, A. N. and Baldwin, T. O. (1997) Cotranslational protein folding. *J. Biol. Chem.* 272, 32715-32718.
22. Kudlicki, W., Chirgwin, J., Kramer, G., and Hardesty, B. (1995) Folding of an enzyme into an active confirmation while bound as a peptidyl-tRNA to the ribosome. *Biochemistry* 34, 14284-14287.
23. Makeyev, E. V., Kolb, V. A., and Spirin, A. S. (1996) Enzymatic activity of the ribosome-bound nascent polypeptide. *FEBS Lett.* 378, 166-170.
24. Schaffitzel, C., Hanes, J., Jermutus, L., and Plückthun, A. (1999) Ribosome display: an *in vitro* method for selection and evolution of antibodies from libraries. *J. Immunol. Methods* 231, 119-135.
25. Wilson, D. S., Keefe, A. D., and Szostak, J. W. (2001) The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. USA* 98, 3750-3755.
26. Doi, N. and Yanagawa, H. (1999) STABLE: protein-DNA fusion system for screening of combinatorial protein libraries *in vitro*. *FEBS Lett.* 457, 227-230.
27. Cheadle, C., Ivashchenko, Y., South, V., et al. (1994) Identification of a Src SH3 domain binding motif by screening a random phage display library. *J. Biol. Chem.* 269, 24034-24039.
28. Gram, H., Schmitz, R., Zuber, J. F., and Baumann, G. (1997) Identification of phosphopeptide ligands for the Src-homology 2 (SH2) domain of Grb2 by phage display. *Eur. J. Biochem.* 246, 633-637.
29. Frankel, A., Welsh, P., Richardson, J., and Robertus, J. D. (1990) Role of arginine 180 and glutamic acid 177 of ricin toxin a chain in enzymatic inactivation of ribosomes. *Mol. Cell Biol.* 10, 6257-6263.
30. Kim, Y., Mlsa, D., Monzingo, A. F., Ready, M. P., Frankel, A., and Robertus, J. D. (1992) Structure of a ricin mutant showing rescue of activity by a noncatalytic residue. *Biochemistry* 31, 3294-3296.
31. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
32. Hanes, J., Jermutus, L., and Plücktun, A. (2000) Selecting and evolving functional proteins *in vitro* by ribosome display. *Methods Enzymol.* 328, 404-430.
33. Sawata, Y. S., Wada, A., and Taira, K. (2003) An advanced ribosome-display with strengthened association (*ARiSA*) for *in vitro* selection of a peptide aptamer with strong affinity. Manuscript in preparation.
34. Sawata, S. Y., and Taira, K. (2003) Modified peptide selection *in vitro* by introduction of a protein-RNA interaction. *Protein Eng.* 16, 1115-1124.
35. Sawata, S. Y., Suyama, E., and Taira, K. (2004) A system based on specific protein-RNA interactions for analysis of target protein-protein interactions *in vitro*: successful selection of membrane-bound Bak-Bcl-xL proteins *in vitro*. *Protein Eng. Des. Sel.* 17, 501-508.

14 分隔式自我复制：聚合酶和其他酶的定向进化的一个新方法

Farid J. Ghadessy and Philipp Holliger

概要 分隔式自我复制（CSR）是蛋白酶，特别是聚合酶，定向进化的一种新方法。在 CSR 最简单形式中，它包括一个简单的环路反馈，即聚合酶只复制编码自身的基因（自我复制）。自我复制发生在一个热稳定的油包水乳化液形成的离散、分离、没有交互的空间中。分隔保证了表型和基因型之间的正确联系（也就是说，分隔保证了聚合酶只复制编码自身的基因而与空间外的成分没有关系）。这样，聚合酶特性的改进直接转化到编码该聚合酶的基因信息的扩增上。CSR 被证明在聚合酶的定向进化上是一个很有用的方法。在本章，我们描述了一些有用的 CSR 的实验方法，这些方法成功地利用一个 *Taq* 聚合酶随机突变库，使 *T. aquaticus* Pol I（*Taq*）聚合酶性质得到了改变，如增加了热稳定性以及对潜在抑制剂（肝素）的抵抗能力。

关键词 分隔式自我复制；直接进化；体外筛选；乳化液；*Taq* 聚合酶

14.1 引　　言

自我复制，忠实地拷贝自身的基因组，这是所有生命的一个确定的特性。在当今的生命体中，这个最基本的生命活动由不同种类的多核苷酸聚合酶来实现[1]。除了基因组的扩增外，多核苷酸聚合酶还有一系列核心的生物功能，包括转录、DNA 修复以及端粒维持。近年来，人们发现某些特殊的聚合酶参与很多不同的生命活动，包括从适应突变、抗体亲和力成熟到抵抗紫外线引起的 DNA 损伤[2]，而且，还有可能参与 RNA 干扰[3]。异常的聚合酶功能可引起癌症的发生[4]，许多聚合酶，尤其是病毒的聚合酶，是重要的药物靶标。最后，聚合酶处于现代生物技术发展的中心位置，它用于 DNA 测序、聚合酶链反应、定点突变、DNA 克隆；而且聚合酶还主要应用于新兴技术，如分子计算学和纳米生物技术。我们相信对聚合酶功能理解得更全面，不仅可以帮助我们更深入理解一些生命活动的机制，而且还能使其更广泛用于生物技术领域，还有可能加速设计出抗病毒的药物。

虽然 T. Steitz 和其他科学家对聚合酶进行了三维结构上的研究，而且人们对其生化方面以及野生型和突变型的酶动力学方面也都有了细致的研究，但是在我们想按自己的意愿改变聚合酶的性质或为了新的用途设计一些聚合酶这些方面，其发展仍然很滞后[5]。

14.1.1 通过蛋白质设计实现聚合酶工程技术

通过蛋白质工程技术，可以改变聚合酶的功能特性。例如，通过删除全部或部分

的 *Taq* 聚合酶上的 5′→3′核酸外切酶结构域，得到变异的 *Taq* 聚合酶（Stoffel fragment 和 Klentaq）；该变异的聚合酶增加了热稳定和 DNA 复制的保真度，但 DNA 持续合成的能力降低[6,7]。进一步，高分辨率的聚合酶结构的获得可以通过合理设计一些改进聚合酶性质的突变体来实现（如在 DNA 循环测序中，*Taq* 的突变体增强了双脱氧核酸的整合；参考文献［8］）。定点突变也可以得到聚合酶的变异体，从而使聚合酶增加了整合核苷酸的能力[9,10]、减少了 DNA 复制的中止[11]以及改变了 DNA 复制的保真性[12]。将 T7 DNA 聚合酶的硫氧还蛋白结合环移植到 *Escherichia coli* DNA 聚合酶Ⅰ上，可以显著增强复制的持续性[13]。

14.1.2 通过功能筛选实现聚合酶工程技术

遗传学的方法也已应用到聚合酶的设计中。例如，Loeb 和他的同事通过 *polA*12、*recA*718 菌株与 *Taq* 聚合酶[14]、人类免疫缺陷病毒反转录酶[15]和人类 pol β[16]的互补实验，选出了有活性的变异聚合酶。互补实验筛选被用来检测聚合酶活性位点的可变性，并且用来筛选出有不同特性的聚合酶突变体，如选出保真度降低的突变体或整合核苷酸能力增强的突变体。最近，这种方法被用在 *RAD*30、*RAD*52 酵母株中对 pol η 功能的筛选上，并找到了 pol η 活性增强的突变体[17]。

噬菌斑展示技术在功能筛选方面是个很成功的方法，尤其是该方法可应用在分子相互作用的定向进化方面。例如，该方法在分离肽激素模拟物和人类 V-基因特异的抗体筛选中的应用。近来，噬菌斑展示技术通过在噬菌斑中展示底物和酶，可以用来选择酶的不同催化活性。这种方法被 Jestin 和 Winter 成功用来富集有活性的聚合酶，通过在模板-引物双链体中加入有标签的核苷酸到噬菌体颗粒中[18]。不久前，Romesberg 和他的合作者采用类似的方法对 Stoffel 基因进行了筛选[19]，筛选出的 Stoffel 某个片段整合 rNTP 的效率接近野生型的酶。

虽然遗传互补方法严重的限制了该方法的使用范围，但噬菌斑展示技术显示出了多方面的适用性。但是，噬菌斑展示技术筛选条件需要受噬菌体存活条件的影响，而且噬菌体内在的分子可以降低聚合酶与模板-引物二链体的亲和性以及聚合酶复制的持续性。这两种方法都有能力检测出很弱的聚合酶活性，仅仅一个 dNTP 整合也能被筛选出来。

14.1.3 通过分隔式自我复制实现聚合酶工程技术

我们在酶的进化，特别是聚合酶的进化方面，发明了一种新方法，称为分隔式自我复制（CSR，参考文献［20］）。在 CSR 中，每个不同的聚合酶突变体在不同的隔离空间。使用合适的试剂，使每个聚合酶只复制编码自身的基因，而不受外界的影响（也就是说，自我复制；图 14.1）。相应地，只有编码有活性的聚合酶的基因才能复制；失活的聚合酶不能复制，从而在基因池中消失。不同活性的突变体，活性越高的突变产生越多的自身基因的拷贝数（即产生更多的“子代”基因）。结果，聚合酶在复制后，其拷贝数反映了在给定的条件下其聚合酶活性的高低。这样，最适应选择条件的聚合酶有最大的活性，从而在基因池中占有最主要的地位。

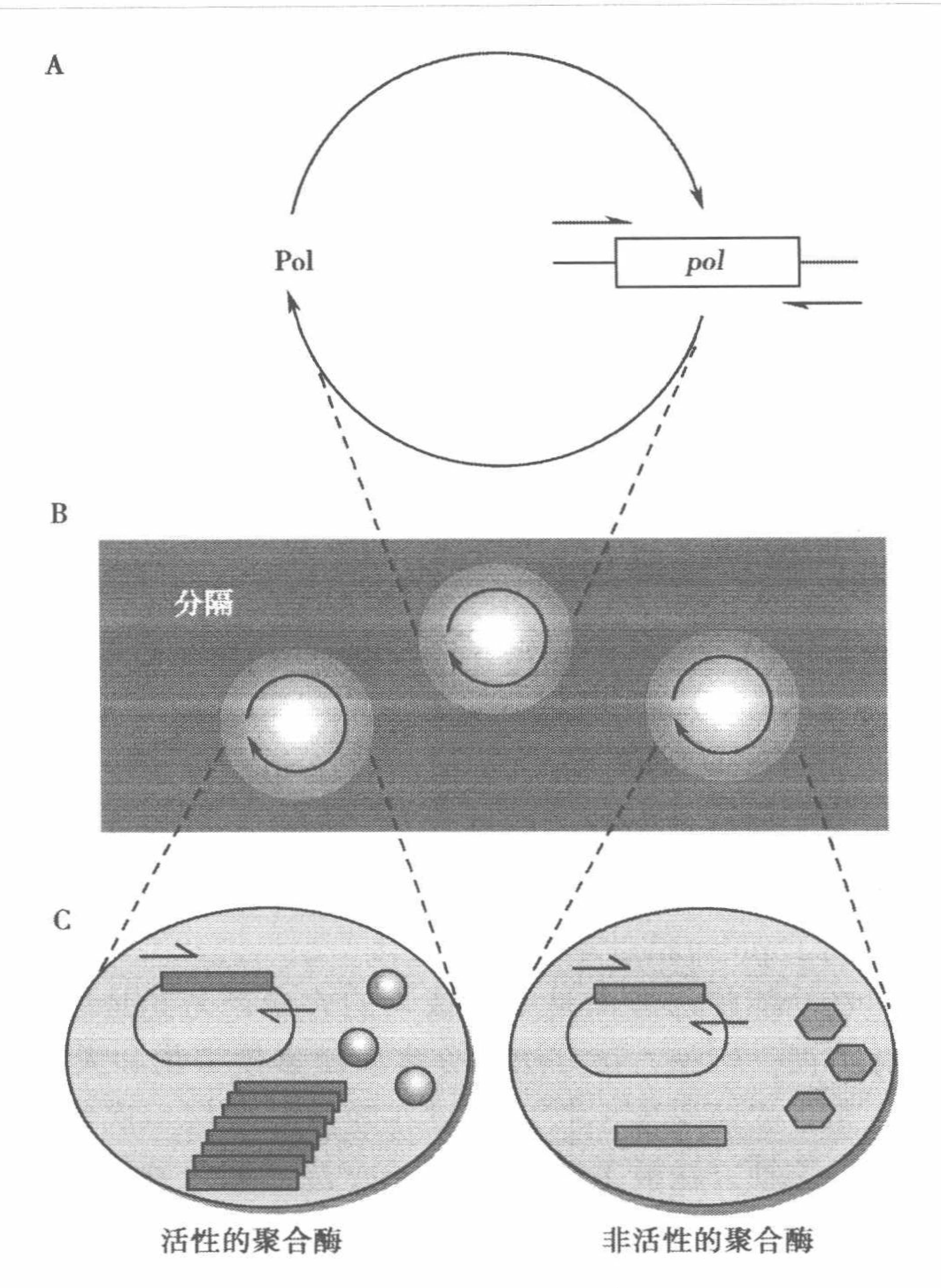


图 14.1　A. CSR包括一个简单的环路反馈，只含有复制编码自己基因的聚合酶。B. 分隔保证了自我复制的反应互不影响。C. 在CSR体系中，酶性质的改变直接（或按比例）地转换到编码该酶的基因扩增的产量上。也就是说，只有活性的聚合酶（深灰色球型）可以复制，失活的聚合酶（浅灰色六边形）不能扩增自身的基因，从而在基因池中消失。

物理上分离的隔室，主要保证在 CSR 中表型和基因型正确对应关系（即保证每个聚合酶特性的改变，只能反映在其复制自身基因的行为上）。分隔空间的完整性在选择的进程中不能破坏，一些来自嗜热物种的聚合酶，如 *Taq* 酶，需要长时间地在高温下，也需要保证分隔空间的完整性。其次，分隔室不能有大分子之间的交换，如 DNA 或蛋白质（小分子的交换是允许的，在某些情况下，还是有利的）。许多不同的方法用来保证酶反应的空间封闭性，如脂质囊泡[21]。我们使用的是油包水乳化剂[22]。我们改进了表面活性剂的成分和油水相的比例，增加了其热稳定性和 CSR 产物的生成。乳化剂在温度超过 90℃时，稳定了体系，而且乳化-PCR 反应与溶液-PCR 反应的产物量基本接近[20]

CSR 乳化剂的高稳定性，可允许其在更宽广的实验条件下进行蛋白酶活性的选择。CSR 可严格选择聚合酶的活性：在一个筛选中，*Taq* 野生型聚合酶稀释到 $1/10^6$，其活性比失活的 *Taq* 突变体（D785H/E786V）还高。这些使得 CSR 成为聚合酶定向进化的一个强有力的方法。实际上，以两个 *Taq* 基因的随机突变库为基础，每个仅仅用

CSR 进行 3 个回合的试验，我们就得到了增强 11 倍热稳定性的 *Taq* 聚合酶突变体；还得到了增强 130 倍的对肝素有抵抗的 *Taq* 聚合酶突变体，而肝素是一个对所有聚合酶都有很强的抑制性，并广泛使用的抗凝血剂[20]。

我们希望 CSR 和其他方法在聚合酶的定性选择方面有更多的应用，包括设计一些聚合酶，使其满足更多的应用领域。例如，染料中止测序、临床或法医诊断 PCR、突变试验等。聚合酶的稳定性、对抑制剂的抵抗性、复制持续性、保真性、对模板-引物双链体的亲和性、底物特异性等都是可以改变的，聚合酶可以在已有应用领域上进行扩展。例如，聚合酶对底物的特异性的改变，可以允许其结合更多的遗传字母，允许核酶和脱氧核酶的体外进化，而且可以帮助改进快速单分子测序的方法。最后，聚合酶不仅可以扩展自身底物的特异性，而且可以改变聚合酶本身的化学性质，这使其可以创造出与目标序列类似的 DNA-like 的聚合体，这样就可以按照人为的意愿进行 DNA 复制。从而，可以将分子进化领域扩展到材料科学领域。

自我复制和分隔也被用来重现分子自组织和生物前进化中的情景。特别地，一种巧妙的选择策略被用来重现 RNA-based 的 RNA 复制酶进行的自我复制。Bartel 和 Joy 的实验组已经成功地创造出不同的核酶，这些核酶可以直接在模板上加入核苷酸[23,24]。一旦 RNA 复制酶被创造，分隔式自我复制（或其他方法）仅仅需要保证复制的是自身的 RNA[25]，阻止其他污染的 RNA 进入体外复制系统。

最后，CSR 不仅应用在聚合酶上，还可以经过催化网络应用于其他酶上。我们将 CSR 应用于核苷酸二磷酸激酶（NDK）的筛选上，NDK 和 *Taq* 聚合酶可完成一对互逆的催化反应，NDK 可生成三磷酸核苷酸，三磷酸核苷酸是聚合酶复制 *ndk* 基因所必需的[20]。结果，只能有活性的 *ndk* 基因可以复制。CSR 还可以整合单酶或反应通路的进化。基于 CSR 的“聚合酶工程”，一个通用的选择酶活的系统试图被建立起来，希望将酶活反应与聚合酶的底物或抑制剂联系起来。因此，只要合适的底物能在 CSR 基本的回路中，CSR 就可用于对任何酶的筛选。

总之，CSR 是一个强有力的新的用来筛选酶活性的方法，我们希望 CSR 在酶的定向进化，特别是聚合酶的定向进化方面有更多的应用，以及在体外进化和分子协同方面的遗传学研究上也有应用。这里，我们详细地描述了以 *E. coli* 作为表达菌株，*Taq* 聚合酶作为靶蛋白，用 CSR 来选择新特性聚合酶的实验流程。

14.2 材 料

（1）*E. coli* 抑制菌株 TG1（K12、Δ［*lac-pro*］、*sup*E、*thi*、*hsd*D5/F′ *tra*D36、*pro*A^+ B^+、*lac*Iq、*lac*ZΔM15）用来扩增质粒和聚合酶的表达。

（2）2×TY 培养基（1L；参考文献［26］）：16g 胰蛋白胨、10g 酵母提取物和 5g 氯化钠；以及 0.1mg/ml 氨苄青霉素（Amp；2×TYA）。

（3）TYE 平板（1L；参考文献［26］）：10g 胰蛋白胨、5g 酵母提取物、8g NaCl 和 15g 氯化钠；以及 0.1mg/ml 氨苄青霉素。

（4）脱水四环素（ACROS）。

（5）10×Super *Taq* 聚合酶缓冲液（HT Biotech，Cambridge，UK）。

（6）四甲基氯化铵（TMAC；Sigma）。

(7) 不含 DNA 酶的胰腺 RNA 酶（Roche）。

(8) CSR 油相：4.5%（*V*/*V*）Span 80（Fluka）、0.4%（*V*/*V*）Tween-80（Sigma）和 0.05%（*V*/*V*）Triton X-100（Sigma）在轻矿物油中（Sigma；如何准备参见 14.3.3）。

(9) 2ml 圆底生物冷冻瓶（Costar，Cambridge MA）。

(10) 磁力搅拌器。

(11) 薄壁 PCR 管（Roche）。

(12) 二乙醚（Sigma）。

(13) 聚乙二醇（PEG）800/$MgCl_2$ 溶液：30%（*V*/*V*）PEG800（BDH Biochemical）和 30mmol/L $MgCl_2$。

(14) Tris-HCl-EDTA（TE）缓冲液：10mmol/L Tris-HCl、pH8.0 和 1mmol/L EDTA。

(15) 20U/μl *Dpn*I（New England Biolabs）。

(16) PCR 纯化试剂盒 Qiagen。

(17) EB（Qiagen）：10mmol/L Tris-HCl，pH8.0。

(18) ExoSap（USB）。

14.3 方　　法

下面的实验流程详细描述用 CSR 在 *E. coli* 中对 *Taq* 聚合酶突变体的筛选。这个方法保证在标准的 PCR 反应条件下，聚合酶突变体是有活性的。用来产生文库的方法是核苷酸类似物[27]和易错 PCR[28]，这里不再描述。引物 1（5′-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GCA A-3′）针对表达载体 pASK75[29]是特异的，用 *Xba*I 和 *Sal*I 将 *Taq* 基因连接到载体 pASK75。引物 2（5′-GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GCG-3′）针对表达载体 pASK75 是特异的。每个引物含有 5′的突起黏末端，这样可以用引物 3（5′-CAG GAA ACAGCT ATG AC-3′）和 4（5′-GTA AAA CGA CGG CCA GT-3′）来加强 CSR 的筛选。

14.3.1 用 Tet 启动子在细菌中表达 *Taq* 聚合酶

(1) 挑取单克隆到 2ml 的 2×TY [见 14.2 (2)] 培养基中，并加入 0.1mg/ml Amp，在 37℃过夜培养。

(2) 将上述培养液以 1∶100 的比例加入到新鲜的 2×TYA 37℃培养，直到菌液 OD_{600} 达到 0.6～0.9（培养应不超过 2～3h）。

(3) 加入 ACROS（非毒性四环素衍生物），终浓度为 0.2μg/ml 到步骤（2）培养液中，诱导 pASK75 种的 Tet 启动子，表达蛋白质。

(4) 在 37℃持续摇菌 2～4h。

14.3.2 收集表达了聚合酶的细胞

(1) 在离心机中 [1300rcf（3000r/min），15min] 离心细胞，去掉上清液，用与培养基相同体积的 1× *Taq* 缓冲液轻轻地悬浮细胞。

(2) 重复离心收集细胞和重悬细胞，这次重悬细胞使用 1/10 培养基体积的 1× *Taq* 缓冲液。

(3) 检测 OD_{600} 的吸收值，确定细胞的个数（OD_{600} =1 相当于 8×10^8 个细胞/ml）。

14.3.3 建立 CSR

（1）将水相的 CSR 在冰上加入到 1× *Taq* 缓冲液，加入 1μmol/L 引物 1 和 2、0.25mmol/L dNTP、50μmol/L TMAC（可选；见注 1）、0.05%（*V/V*）不含脱氧核糖核酸酶的胰腺核糖核酸酶（可选）和诱导表达的细胞（终浓度为 1×10^9 个细胞/ml）。

（2）准备油相 CSR，混合轻矿物油、4.5%（*V/V*）Span 80、0.4%（*V/V*）Tween-80、0.05%（*V/V*）Triton X-100，在室温下持续搅拌。由于表面剂的高黏性，需要准备一个大体积油相 CSR（大于 50ml），并且用切断吸管尖去掉表面活性物质。油相的 CSR 一旦准备好，可以在黑暗的室温中存放 1 个月。

（3）将 200μl 的水相 CSR 加入到 400μl 油相 CSR，加入过程中要用点滴的方法（每滴 5～10μl，间隔 5s），整个过程在 2ml 的平底管中进行，需要用磁力搅拌器（1000r/min）不停地搅拌。

（4）加入最后 1 滴后，持续搅拌 5min。最后需要乳化成白色的并且有黏性的状态（见注 2）。

14.3.4 CSR

（1）转移乳化液到 0.5ml 薄壁 PCR 管（100μl/管），然后加入 2 滴矿物油，防止挥发（见注 3）。

（2）进行 PCR 反应，94℃ 5min，裂解细胞，破坏其他酶的活性，然后进行如下 20 个循环：94℃（1min），60℃（1min），72℃（5min）。

PCR 完毕后，乳化相应该仍是白色的，上面覆盖了一层油相。体积可能会减少，但不表明乳化的隔室发生合并。乳化相的聚合或破碎直接表现为白色的乳化相之下有个明显的水相层。这时候就表明有很多因素干扰了乳化相的稳定，如各种成分混合的不充分。

14.3.5 实施

（1）用 2 倍体积（200μl）的二乙醚，来分离水相的 CSR。漩涡振荡混合物 20s，然后在 20 000rcf（13 000r/min）离心 2min。

（2）出现两个液体层。小心地移动有机相下面的水相层，然后转移到一个新的 1.5ml EP 管中。

CSR 的反应可以通过两种不同的方法来实施。

14.3.5.1 PEG 抽提方法

（1）用苯酚-氯仿（1/1，*V/V*）来萃取液相层，然后接着用氯仿-异戊醇（24/1，*V/V*）来萃取。

（2）在液相层，加入 0.5 个体积的 PEG800/$MgCl_2$ 溶液［见 14.2（13）］，然后用吸量管上下反复混匀几次。

（3）室温，20 000rcf（13 000r/min）离心 10min。去掉上清液（含有剩余的引物和 dNTP），沉淀用 TE 重悬［见 14.2（14）］。

（4）遗留的聚合酶，在加入 20～50U *Dpn*I 到液相层中，37℃放置 1h 后，其量可

减少（见注 4）。

（5）为了确保完全去除引物，进一步利用 PCR 纯化试剂盒（Qiagen），纯化液相层中的 CSR 产物。用 35%（*V*/*V*）盐酸胍洗 2 次，保证完全去除残留的引物。最后将纯化的产物加到 50μl 的缓冲液 EB 中［Qiagen；见 14.2（16）］。

14.3.5.2　快速实施

（1）利用 14.3.5.1 步骤（5）中的方法，用 PCR 纯化试剂盒（Qiagen）纯化 CSR 产物。用 35%（*V*/*V*）盐酸胍洗 2 次，保证完全去除残留引物。最后将纯化的产物加到 50μl 的缓冲液 EB 中。

（2）取 7μl 的产物，加入 1μl 的 10× *Taq* 缓冲液、1μl 的 20U/μl *Dpn*Ⅰ以及 1μl ExoSAP，然后在 37℃放置 1h，接着在 85℃放置 15min。

14.3.6　提取

（1）使用外侧巢式引物 3 和 4，2 次扩增需要筛选的基因。20%的需要筛选的产物被加入 50μl 的 PCR 反应液中（即 10μl 的纯化产物加入到 50μl 溶液中）。但是，这个用量有可能增多或减少，需要根据第一次的结果（见注），2 次扩增的反应条件需要根据第一次扩增的结果来调整，常用的是 25 个循环：94℃（30s）、50℃（30s）、72℃（5min）。循环的次数也可以根据结果来调整（见注）。

（2）结果用琼脂糖凝胶来检测。成功的筛选是出现一条正确大小的条带（对于聚合酶，是 2.5kb），而在负对照中不应出现条带［见注 5 和注 6］。

CSR 的条带可进一步切胶纯化，再次酶切，重新克隆到载体 pASK75 上。然后转化到 TG1［见 14.2（1）］，转化的产物或者按 14.3.7 所描述的方法筛选，或者按 14.3.2～14.3.6 中所描述的方法培养、诱导，为下一轮的选择做准备。

14.3.7　*Taq* 聚合酶的快速筛选

（1）为了筛选突变体，转化的克隆在 TYE-Amp 平板上扩增［见 14.2（3）］，然后接种到 96 孔板上，每孔含有 100μl 的 2×TYA［见 14.2（2）］。

（2）上述 96 孔板培养 6～10h 以后，即加入 20μl 含有 1.2μg/ml ACROS 的 2×TYA，继续培养 2～4h，表达蛋白质；

（3）用排枪，从 96 孔的每个孔转移 2μl 诱导的细胞到 30μl 的 PCR 混合物中（见 14.3.3 和 14.3.4），该混合物也在一个 96 孔的 PCR 平板上。

（4）在 96 孔 PCR 平板中，每孔加入覆盖油，然后进行 PCR 反应［见 14.3.4（2）］，可以将循环增加到 30 次，用来筛选活性比较弱的突变体。用野生型的作为对照很重要。阳性鉴定可在琼脂糖凝胶上进行。

14.4　注

1. TMAC 用来增加引物的特异性，在某些情况下不是必需的。

2. 为了测试新准备的油相，标准的 PCR 反应可使用 14.3.3 和 14.3.4 中的实验流程。在乳化相中的 PCR 反应产物的量应当与非乳化相 PCR 反应的产物差不多。在

PCR 反应过程中，乳化相不能有明显的破坏，一般根据是否在 PCR 管底部形成一个清晰的水相层来判断。

3. 如果 PCR 反应没有使用油覆盖或是使用了加热盖，CSR 的产量会变少。

4. 甲基化的质粒 DNA 在筛选的过程中可以作为非特异的模板，所以需要被去除。非甲基化的扩增产品不需要被去除。

5. 在所有的 CSR 筛选中，阴性对照是很重要的。不加入 dNTP 的 CSR 反应，是一个理想的负对照。通过重复扩增，在负对照中，不应出现可见的扩增产物。若是出现筛选产物，表明没有完全去除引物和质粒。

6. 如果两次扩增出现了很弱的信号，那么外侧巢式引物 3 或 4 需要与基因特异性的引物一起使用，这样有时可以增强特异的信号。在任何情况下，包括－dNTP 的负对照都是很重要的。

参考文献

1. Steitz, T. A. (1999) DNA polymerases: structural diversity and common mecha-nisms. *J. Biol. Chem.* 274, 17, 395-17, 398.
2. Goodman, M. F (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* 71, 17-50.
3. Nishikura, K. (2001) A short primer on RNAi: RNA-directed RNA polymerase acts as a key catalyst. *Cell* 107, 415-418.
4. Goldsby, R. E., Lawrence, N. A., Hays, L. E., et al. (2001) Defective DNA polymerasedelta proofreading causes cancer susceptibility in mice. *Nat. Med.* 7, 638, 639.
5. Joyce, C. M. and Steitz, T. A. (1994) Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63, 777-822.
6. Barnes, W. M. (1992) The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* 112, 29-35.
7. Lawyer, F. C., Stoffel, S., Saiki, R. K., et al. (1993) High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5′ to 3′ exonuclease activity. *PCR Meth. Appl.* 2, 275-287.
8. Li, Y., Mitaxov, V., and Waksman, G. (1999) Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. USA* 96, 9491-9496.
9. Patel, P. H. and Loeb, L. A. (2000) Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.* 275, 40266-40272.
10. Astatke, M., Ng, K., Grindley, N. D., and Joyce, C. M. (1998) A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl. Acad. Sci. USA* 95, 3402-3407.
11. Ignatov, K. B., Bashirova, A. A., Miroshnikov, A. I., and Kramarov, V. M. (1999) Mutation S543N in the thumb subdomain of the *Taq* DNA polymerase large fragment suppresses pausing associated with the template structure. *FEBS Lett.* 448, 145-148.
12. Kunkel, T. A. and Bebenek, K. (2000) DNA replication fidelity. *Annu. Rev. Biochem.* 69, 497-529.
13. Bedford, E., Tabor, S., and Richardson, C. C. (1997) The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I. *Proc. Natl. Acad. Sci. USA* 94, 479-484.
14. Suzuki, M., Baskin, D., Hood, L., and Loeb, L. A. (1996) Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites *in vivo* with the crystal structure. *Proc. Natl. Acad. Sci. USA* 93, 9670-9675.

15. Kim, B., Hathaway, T. R., and Loeb, L. A. (1996) Human immunodeficiency virus reverse transcriptase. Functional mutants obtained by random mutagenesis coupled with genetic selection in *Escherichia coli*. *J. Biol. Chem.* 271, 4872-4878.

16. Sweasy, J. B. and Loeb, L. A. (1993) Detection and characterization of mammalian DNA polymerase beta mutants by functional complementation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 90, 4626-4630.

17. Glick, E., Vigna, K. L., and Loeb, L. A. (2001) Mutations in human DNA polymerase eta motif II alter bypass of DNA lesions. *EMBO J.* 20, 7303-7312.

18. Jestin, J. L., Kristensen, P., and Winter, G. (1999) A method for the selection of catalytic activity using phage display and proximity coupling. *Angew. Chem. Int. Ed.* 38, 1124-1127.

19. Xia, G., Chen, L., Sera, T., Fa, M., Schultz, P. G., and Romesberg, F. E. (2002) Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. USA* 99, 6597-6602.

20. Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001) Directed evolution of poly-merase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. USA* 98, 4552-4557.

21. Oberholzer, T., Albrizio, M., and Luisi, P. L. (1995) Polymerase chain reaction in liposomes. *Chem. Biol.* 2, 677-682.

22. Tawfik, D. S. and Griffiths, A. D. (1998) Man-made cell-like compartments for molecular evolution. *Nature Biotechnol.* 16, 652-656.

23. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* 292, 1319-1325.

24. McGinness, K. E., Wright, M. C., and Joyce, G. F. (2002) Continuous *in vitro* evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem. Biol.* 9, 585-596.

25. Szostak, J. W., Bartel, D. P., and Luisi, P. L. (2001) Synthesizing life. *Nature* 409, 387-390.

26. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1990) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

27. Zaccolo, M. and Gherardi, E. (1999) The effect of high-frequency random mutage-nesis on *in vitro* protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol.* 285, 775-783.

28. Vartanian, J. P., Henry, M., and Wain-Hobson, S. (1996) Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res.* 24, 2627-2631.

29. Skerra, A. (1994) Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene* 151, 131-135.

15 Raf 蛋白 Ras 结合结构域的简并进化库合成以及利用片段互补法快速筛选二氢叶酸还原酶的快速折叠且稳定的克隆

Francois-Xavier，Campbell-Valois and
Stephen W. Michnick

概要 蛋白质工程领域主要涉及关于设计新的酶活性或折叠，以及理解蛋白质正确折叠和稳定的最基本的序列决定因素。迄今，已有巨大的精力投入到设计构建多肽文库方法的研究中。最常用的方法是用来筛选具有高亲和力的特异性结合蛋白的噬菌体展示技术。本章讲述了一种可替代噬菌体展示的方法，该方法是基于二氢叶酸还原酶（DHFR）蛋白质片段互补检测（PCA）建立起来的，可以完全用于体内实验。我们以筛选 *raf* 蛋白的 *ras* 结合结构域（*ras*-binding domain，RBD）正确折叠且能与 ras 结合所需的简并序列为例，讲述 DHFR PCA 的应用。此筛选系统通过重复的竞争实验，富集文库中表现最好的序列，而无需将体外实验所必需的文库筛选和扩增步骤分离。而且，可以在 96 孔板中通过三乙酸基氮（nitrilotriacetic acd，Ni-NTA）亲和层析直接快速处理被选中的克隆。此方法特别适用于设计和筛选研究序列折叠和结合决定因素的文库。另外，将该方法略加修改，还可以应用于文库间的筛选，从而可以使相互作用的蛋白质实现共同进化。

关键词 蛋白质片段互补检测；二氢叶酸还原酶；细菌存活检测；噬菌体展示；蛋白质相互作用；蛋白质工程；蛋白质折叠；简并进化库；聚合酶链反应；结合检测；6×His-tag 亲和纯化

15.1 引　　言

随着 20 世纪七八十年代重组 DNA 技术的出现，已发展出大量的合成和筛选寡核苷酸文库的新方法，用于研究和了解是哪些序列编码了具有期望特性的新蛋白质，这些特性包括在特定条件下酶的活性、结合力以及蛋白质的稳定性[1~8]。对于任何一个给定的例子，要进行此类蛋白质的工程研究，必须具备两个条件：构建多样化序列文库的策略和从文库中筛选具有特定性质产物的有效方法。其中，选择何种方法合成文库至关重要，因为这决定了文库能否提供大量充足的筛选空间，从而最大限度的为得到编码蛋白质特定性质的序列提供最多的选择。因此，发展构建文库的策略毫无疑问一直是这个领域的研究焦点。过去的文献，报道了一些将蛋白质中的某区域完全随机化或高度简并化的例子，以回答关于蛋白质折叠的问题[9,10]。然而，关于真正建立高度简并文库，寻找编码新的折叠、结合能力或酶活的序列的报道仍很少见[6,11]。构建高度随

机文库固有的以及随后在大量的序列库中寻找编码特定性质的少数序列的局限性，推进了新方法的设计和开发，这其中就包括 DNA 混编（DNA shuffling）策略或允许毫无序列相似性的基因间重组的新方法[12～18]。

如果没有适当的筛选策略，即使文库设计得再精巧也无法得到有用的产物。例如，在一个文库中筛选编码具有特定酶活、稳定性或具有结合靶标蛋白能力的蛋白质的理想方法是，在某细胞或有机体中表达该文库，而文库中具有所需性质的成员可以在特定的条件或基因背景下，在特定的选择培养基中生长[1,19～22]。但是，这种例子是极为少见的，所以从事蛋白质工程研究的人员一直在寻求可以更为普遍应用的文库筛选方法[23～26]。更具体地说，结合检测经常可以用于在一个 DNA 文库中筛选表达正确折叠的、稳定的和可以与其他分子（蛋白质、核酸、有机底物、过渡态类似物等）结合的多肽。目前建立的最完善的方法是噬菌体展示[4,5,27]；综述[28～30]，该方法将文库的扩增和筛选分开，分别在体内和体外进行，适用于与可以容易地固定在固体介质上的小分子或小肽结合的蛋白质，但不太适用于蛋白质与蛋白质相互作用研究或文库间筛选。近来，蛋白质片段互补检测（PCA）成为一个替代选择[31～34]。PCA 的策略是，将两个蛋白质分别与报告蛋白或酶的两个片段融合表达，两个目的蛋白的结合将报告蛋白的两个片段连在一起，从而使报告蛋白具有可以检测的功能。例如，以鼠二氢叶酸还原酶（mDHFR）作为报告蛋白，在大肠杆菌中利用简单的存活选择实验进行文库筛选的方法已经建立起来[32]。与在所有的原核和真核细胞中相同，在大肠杆菌中，DHFR 的产物四氢叶酸是合成胸腺嘧啶、甘氨酸、丝氨酸和腺嘌呤所必需的，在原核生物中，它还是合成泛酸所必需的。因此，在不提供 DHFR 终产物的条件下，DHFR 的活性显然对于细胞生长和分裂是不可缺少的。用甲氧苄氨嘧啶（trimethoprim，叶酸类似物，其抑制大肠杆菌 DHFR 的能力是抑制哺乳动物 DHFR 的 12 000 倍）处理的大肠杆菌的生长依赖于重组 mDHFR 的表达[35]。这样，将两个目的蛋白分别与 DHFR 的两个片段融合并共表达，如果两个蛋白质之间有相互作用，则大肠杆菌可以在添加了甲氧苄氨嘧啶的基本培养基上生长[31]。在一个最初的文库间筛选中，DHFR PCA 被用于从含有 6×10^{10} 个可能结合的序列的单独文库中鉴定出可以形成最优的形成亮氨酸拉链的异源二聚体序列。竞争实验（competition experiment）和“文库混编”（library shuffling）策略旨在改善文库筛选的覆盖范围，进一步优化形成二聚体的蛋白质对，并最终找到一个“成功的蛋白质对”[32,36]。近来，DHFR PCA 经修改用于在体内筛选单链抗体[37]。DHFR PCA 策略将基因筛选整合的方法，是可以简化文库间筛选的关键，因为在若干的循环中选择压力同时施加于两个文库上，从而避免了如噬菌体展示技术那样将文库扩增和筛选分开进行，这使 DHFR PCA 真正可以用来研究寡聚体序列的同时进化。

亮氨酸拉链例子的结果使我们相信 DHFR PCA 可以用于解决更具挑战性的问题。在此例中，仅仅改变了几个关键氨基酸的位置，而且只有 2～4 个氨基酸可以被替换。基于之前的理论研究[38]，我们试图严格并全面的测定使 Ser/Thr 蛋白激酶 *raf* 的 *ras* 结合结构域（RBD）正确折叠的序列决定因素[39]。同之前发表噬菌体展示策略类似，我们所用方法的目标是鉴定出蛋白质折叠和稳定所必需的序列[40]。其原理是：如果一个给定的 RBD 变体可以快速正确的折叠并足够稳定，那么它应该可以与 *ras* 相互作用。将 RBD 文库和 *ras* 分别与 DHFR 两个互补的片段融合并在大肠杆菌中共表达，在本节

上述的选择压力下，RBD 文库中可以快速折叠的稳定成员将与 *ras* 相互作用，使 DHFR 恢复活性从而使得大肠杆菌可以生长。我们选择了一个高效且现实可行的方法设计文库，最大限度的开发了其序列多样性并将文库限定在合理的大小。基于此研究提出的问题，一个有意义的扩展序列筛选空间的方法是，在构建文库时每次只改变一小段连续的氨基酸残基[10]。经过对 RBD 结构的研究，我们发现可以将其分割成 13 个区域，分别对应多个独立的β转角或环、β链和一个α螺旋（我们构建了两个文库，分别对应该区域的 N 端和 C 端），其长度为 4～8 个残基[41,42]。以此为依据，我们构建了 13 个简并进化库，用 NNK（其中 N 表示任何核苷酸，K 表示 G 或 T）取代野生型蛋白的密码子，这就使得 20 种氨基酸中的任何一种都可以插入到每个可变区。然后，在大肠杆菌中利用 DHFR PCA 筛选这些文库中可以与 *ras* 结合的成员。除了筛选可以完全在体内进行，该方法最大的优点是得到的文库中可以与 *ras* 相互作用的成员无需更换表达系统即可用于纯化和生理分析。

基于在设计和合成文库以及用 DHFR PCA 对文库进行筛选过程中遇到的技术挑战，在此我们介绍了实验方法和问题解决策略。希望这些方法足够全面，不仅对对折叠有兴趣的人有用，而且对解决优化蛋白质之间的相互作用的问题也有帮助。

15.2 材　　料

15.2.1 文库合成

（1）寡核苷酸引物（IDT）：引物中含有可以与多种碱基配对的位点，以确保获得期望的碱基插入。引物是 SDS-PAGE 纯化得到的。

（2）*Taq* 聚合酶（Fermentas）。

（3）琼脂糖凝胶：agarose（Bioshop）Dark reader™（Clare Chemical Research）和 Gelstar™（Biowittaker Molecular Applications）。

（4）凝胶纯化试剂盒：QIAEX™ II 或 QIAquick™ gel extraction kit（Qiagen）更好。

15.2.2 文库的克隆和纯化

（1）质粒 pQE-32ΔF［1，2］（在 Qiagen 的质粒 pEQ-32 的基础上构建而成，F［1，2］代表 DHFR 片段 1）。

（2）质粒 pREP4（携带 *lac* 抑制基因和卡那霉素抗性筛选标记。含有利用质粒 pQE-32 表达的蛋白质的细胞，正如本实验所用的，必须带有此质粒以控制 pQE-32 质粒上 *lac* 启动子的泄漏表达）（Qiagen 产品）。

（3）连接：T4 DNA 连接酶（Fermentas）和 ATP（Pharmacia）。

（4）SS320 电转感受态细胞（见 15.2.7）。

（5）Genepulser™ II 电转仪（Bio-Rad）。

（6）槽宽 2mm 的电转杯（Invitrogen）。

（7）10ml SOC 培养基：1%蛋白胨、0.5%酵母粉、1% NaCl、0.4%葡萄糖、2.5mmol/L KCl 和 10mmol/L $MgCl_2$。

（8）含 10μg/ml 四环素、10μg/ml 壮观霉素（spectinomycin）和 10μg/ml 氨苄青

霉素的 LB 固体培养基。

(9) 每文库 100～250ml LB 培养基。LB 培养基中添加 0.2%葡萄糖、0.25 倍的 M9 盐溶液 [见 15.2.3 (6)]、10μg/ml 四环素、10μg/ml 壮观霉素和 10μg/ml 氨苄青霉素。

(10) Plasmid Midi Kit (12143) (Qiagen)。

15.2.3　文库的筛选

(1) BL21 电转感受态细胞，先转化质粒 pREP4 [见 15.2.2 (2)]，再转化 pQE-32ΔF [3] (F [3] 代表 DHFR 片段 2)。

(2) Genepulser™ II electroporator system (BioRad) 或 electroporator 2510 (Eppendorf)。

(3) 槽宽 1mm 的电转杯 (Invitrogen)。

(4) SOC 培养基 [见 15.2.2 (7)]。

(5) PBS 缓冲液 (1L)：8g NaCl、0.2g KCl、1.44g Na_2HPO_4、0.24g KH_2PO_4，用 HCl 调 pH 至 7.4，高压灭菌。

(6) 15.1 中所述合适抗生素的 M9 基本培养基 (选择培养基)。1L：740ml 2.5% noble agar (Difco)、200ml 5×M9 盐 (1L M9：64g Na_2HPO_4、15g KH_2PO_4、2.5g NaCl、5g NH_4Cl，参考文献 [43])，加入 2ml 1mol/L $MgSO_4$、1ml 100mmol/L $CaCl_2$ 和 20%葡萄糖。所有的盐和葡萄糖都是细胞培养级别，除了 100μg/ml 氨苄青霉素、25μg/ml 卡那霉素、1mmol/L IPTG、10μg/ml 甲氧苄氨嘧啶、800μg/ml 酪蛋白水解物 (casamino acid)、10μg/ml 硫胺素 (thiamine)。所有溶液均须用去离子水配制并抽滤除菌或高压灭菌 (抗生物、酪蛋白水解物、IPTG 和硫胺素－20℃保存，盐溶液室温存放)。注意 $MgSO_4$、$CaCl_2$ 和葡萄糖溶液要分别灭菌。这种重组培养基被倒入直径为 100mm 或 150mm 的培养皿中，在 4℃可存放 2 个月。

(7) 质粒中提试剂盒 (Qiagen) 或碱裂解的质粒大量提取。

(8) 限制性内切核酸酶：*Hpa*I、*Xma*I、*Eco*NI 和 *Xba*I (NEB 或 Fermentas)。

(9) XL-1 Blue 感受态细胞 (chemiocompetent cell) (见 15.2.8)。

15.2.4　克隆竞争实验

(1) 玻璃培养皿或 15ml 锥形管 (Corning)。

(2) 固体和液体选择培养基 [与 15.2.3 (6) 相同，除液体培养基中不加琼脂]。

(3) Plasmid Midi Kit (Qiagen) 或 alkaline lysis maxiprep。

(4) 含 100μg/ml 氨苄青霉素和 25μg/ml 卡那霉素的 LB 培养基。

15.2.5　克隆分离与测序

(1) 限制性内切核酸酶：*Hpa*I、*Xma*I、*Eco*NI 和 *Xba*I (NEB 或 Fermentas)。

(2) XL-1 Blue 感受态细胞。

(3) 24 孔板 (Corning)，含 100μg/ml 氨苄青霉素的 LB 培养基。

(4) 2ml V-shape 96-well culture block (VWR)。

(5) Montage™ Plasmid Miniprep 96 kit (Millipore, LSKP 09601) 或诸如 QIA-

prep™ Spin Miniprep kit（27104）（Qiagen）的小量纯化试剂盒，这取决于样品数量。

（6）Vacuum manifold Multiscreen Resist™（Millipore，MAVM 096 OR）。

（7）质粒文库（IDT）特异测序引物。

15.2.6 蛋白质纯化

（1）合适的限制性内切核酸酶（对这个例子是 *Sal*I 和 *Xho*I）和连接所需试剂（见15.2.2）。

（2）BL21 pREP4 感受态细胞和含 100μg/ml 氨苄青霉素和 25μg/ml 卡那霉素的 LB 培养基。

（3）Terrific broth（TB）培养基（12g 蛋白胨、24g 酵母粉、4ml 甘油、2.31g KH_2PO_4 和 12.54g K_2HPO_4，加水至 1L）添加 100μg/ml 氨苄青霉素和 25μg/ml 卡那霉素。

（4）50ml 锥形管。

（5）可以离心 96 孔板的离心机和转子，如 Eppendorf 5810 或 5810 R 和 A-4-62。

（6）Ni-NTA Spin Kit 或 Ni-NTA Superflow™ 96 Biorobot Kit（Qiagen），这取决于样品数量。我们使用下列较为便宜的产品替代 Superflow™ 96 Biorobot Kit：Ni-NTA Superflow resin（Qiagen）、0.25mm 的 96 孔玻璃纤维过滤板（3510）、0.2μm 的 96 孔 PVDF 膜板（3504）、96 孔 volume extender（3584）和收集器（Corning）（3958）。

（7）Vacuum manifold Multiscreen Resist™ 和一个大型收集器以及密封块（Millipore，MAVM 096 OR 和 OT）。

（8）Buffer A：6mol/L 盐酸胍、0.1mol/L NaH_2PO_4、0.01mol/L Tris-HCl，pH8.0，添加 10 μmol/L 苯甲基磺酰氟、7.2mmol/L β-巯基乙醇、15～25mmol/L 咪唑和 300mmol/L NaCl。

（9）Buffer B：同 Buffer A，但 pH6.3，添加 7.2mmol/L β-巯基乙醇，有时也添加 15mmol/L 咪唑。

（10）Buffer E：4mol/L 盐酸胍和 0.025mol/L NaOAc pH4.5。

（11）DTT。

（12）6mol/L KOH。

15.2.7 制备 SS320 和 BL21 pREP4 pQE-32 ras-F[3] 电转感受态细胞

（1）SS320 和 BL21 pREP4 pQE-32 ras-F 的过夜培养物。

（2）500ml SOB 培养基：2%蛋白胨、0.5%酵母粉、1ml 10mmol/L MaCl、2.5mmol/L KCl、10mmol/L $MgCl_2$ 和 10mmol/L $MgSO_4$ 和 0.2%葡萄糖，用于 SS320 菌株。

（3）500ml 含有 0.2%葡萄糖的 LB 培养基，用于 BL21 pREP4 菌株。

（4）2L 灭菌冰去离子水。

（5）10%灭菌的甘油。

15.2.8 制备 XL-1Blue 和 BL21 pREP4 电转感受态细胞

（1）XL-1Blue 和 BL21 pREP4 过夜培养物。

（2）500ml 含有 0.2%葡萄糖的 SOB 培养基，用于 SS320 菌株。

（3）500ml 含有 0.2%葡萄糖的 LB 培养基，用于 BL21 pREP4 菌株。

(4) 转化缓冲液：10mmol/L Pipes、15mmol/L $CaCl_2$、250mmol/L KCl，用 KOH 调 pH 至 6.7，然后加入 $MnCl_2$ 至终浓度为 55mmol/L。

(5) Dimethylsulfoxide (DMSO)。

15.3 方　　法

15.3.1 概论

15.3.1.1 PCA 对空间排列的要求

PCA 片段的三维朝向对于 PCA 报告蛋白是否能正确折叠至关重要，它取决于形成复合体的目的蛋白 N 端或 C 端的朝向（图 15.1 示意了在连接设计中对空间位置的

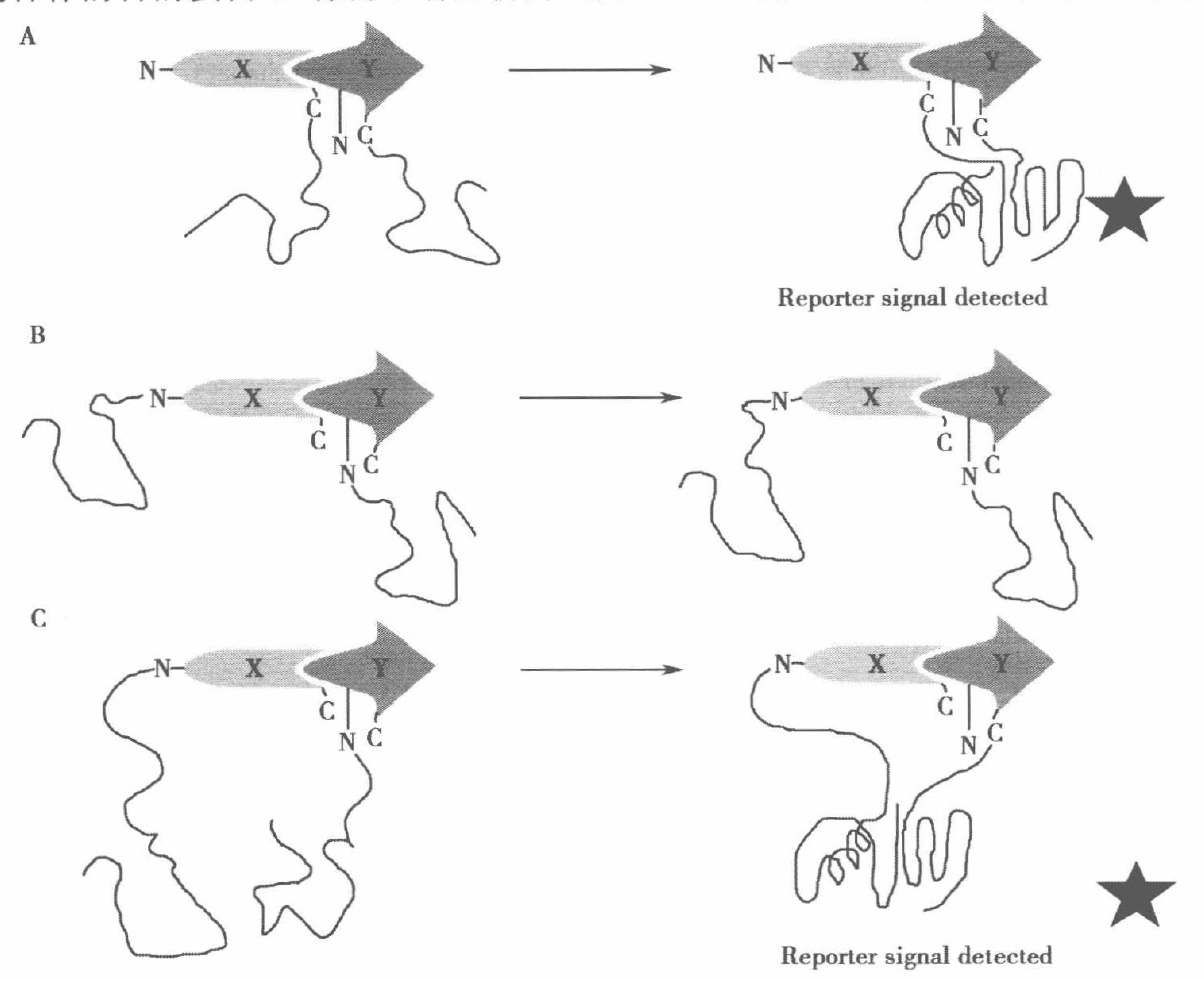


图 15.1　在蛋白质 X 和蛋白质 Y 之间形成二聚体的结构示意图。在所形成的二聚体复合物中，两个蛋白质的 N 端远离，而 C 端虽然在空间上靠近，但却没有直接地接触。在这种情况下，片段可以如 A 中所描述的那样被连接。假定肽键长约 3.75Å，并考虑所研究的 X-Y 复合体在结构和空间上的限制，以及 PCA 报告蛋白在修饰前后的特性，在每一个接头中的氨基酸数目可以被估算出来。如果 X-Y 复合体的结构还没有被测定，接头的长度就只能通过经验来测定了。在有些情况下，可能不能在最佳方向来制备融合。例如，如 A 中所示，X 蛋白的 C 端必须是自由的以便可以结合到 Y 蛋白上。也可以如 B 中所描述的来设计。在这种情况下，很显然如果用 A 中所示的较短的接头，具有天然拓扑特性的 PCA 报告蛋白将不会被重构出来。然而，如果用 C 中所示的较长的接头，就使得从同源片段来的报告蛋白的折叠成为可能。因此，就使得蛋白融合方向充分适合基于 PCA 蛋白工程的筛选策略。

考虑)。因此，在设计蛋白—PCA 片段融合体时，以何种方式将蛋白质片段连接在一起使其能够折叠形成天然结构是非常重要的。融合蛋白能否正确折叠受融合蛋白的末端朝向和片段间接头(linker)长度两个因素的影响。我们将片段 GGGGS 应用于多种不同蛋白质，证明这个序列在大肠杆菌、酵母和哺乳动物细胞表达系统中效果都比较好。我们认为这个序列可以改善融合蛋白的弹性和可溶性，从而有利于其组装，而且由于此序列中不含天然蛋白酶识别位点，确保了融合蛋白的稳定性。虽然此类的接头受到青睐，但并不表示使用它就一定能得到可以互补的片段[31,32,39]。然而，还是要避免使用诸如脯氨酸和β-分支氨基酸类的大的疏水和刚性氨基酸。在大多数蛋白质工程的问题中，形成复合体的目的蛋白的三维结构是已知的，因此可以设计复合体的方向并计算出所需接头的长度。对于 DHFR PCA 来说，融合蛋白 N 端和 C 端的空间构象需求是已知的[31,44]。例如，如果将目的蛋白分别与 F［1，2］的 C 端和 F［3］的 N 端融合，那么构建时所需的接头就可以很短甚至不需要，因为这种构建方法已经考虑到了 DHFR 的拓扑结构；但是如果将目的蛋白分别构建到两个片段的 N 端，要使 DNFR 能正确组装，就必须在两个融合蛋白之间插入至少两个氨基酸作为接头(每个肽键大约 3.75Å)，因为 DNFR 两个片段 N 端的距离将近 10Å。在 *ras* 和 RBD 的例子中就选择了这种构建方式。然而，观察 *raf* RBD 与 ras 高类似蛋白 *rap1A* 的复合体结构发现这两个蛋白质的 C 端距离 40Å[45]，这就需要在每个融合蛋白中插入至少 6 个氨基酸的接头。对于文库筛选来说，我们将每个接头的总长度定为 14 个氨基酸，其中包括限制性内切核酸酶位点，以确保融合蛋白具有足够的灵活性。

15.3.1.2 对照和严密性

在利用 DHFR PCA 进行蛋白质工程研究和文库筛选之前，必须做严格的对照实验用以估计其对特定的测试系统的灵敏度和严密性。理想条件下，在进行文库筛选前，实验者应大概了解 PCA 对于一个给定的相互作用体的解离常数 K_d 的极限。不同的相互作用蛋白对有着不同的灵敏度极限(PCA 可以检测到的最大 K_d)，但是 K_d 值又受到融合蛋白表达水平、蛋白质表达总量中可溶蛋白比例以及诸如稳定性、可溶性和蛋白质折叠、结合参数等蛋白质自身性质的影响。如果 PCA 十分灵敏，可以检测到两个特定蛋白质间非常弱的相互作用，它就不能在众多的克隆中选出最好的相互作用蛋白对；也就是说，此项实验过于灵敏从而丧失了严密性。因此，平衡灵敏度和严密性的关系至关重要。为了解决这一问题，在 PCA 实验前通常需要做些对照实验，虽然不是所有的这些对照都与特定的蛋白质工程研究相关，我们在后面还会再讨论这个问题。

(1) 伪组装(spurious reassembly)：要使 PCA 能有效工作，应避免片段间微弱的或非特异性的相互作用。对于一个给定的相互作用蛋白对测试系统来说，PCA 有效的灵敏度可以通过对照 4 和 6 来估算。如果灵敏度过高，如长出不该有相互作用的蛋白对的菌落(图 15.2A)，可以通过降低表达水平或通过对照 2 中所述的严密性突变来降低灵敏度(图 15.2B)。

(2) 严密性突变(stringency mutant)：突变 DHFR 两片段相互作用表面上的氨基酸残基侧链的效果，如已报道的 F［3］的 Ile114 突变[31,32]。在克隆表达通过形成亮氨酸拉链(leucin zipper)而相互作用的蛋白对时，菌落生长速度和数量无法区分不同蛋白对间的结合效率。后来，在 F［3］中引入突变 I114A，改变了 PCA 的灵敏度，使其可以应用于亮氨酸

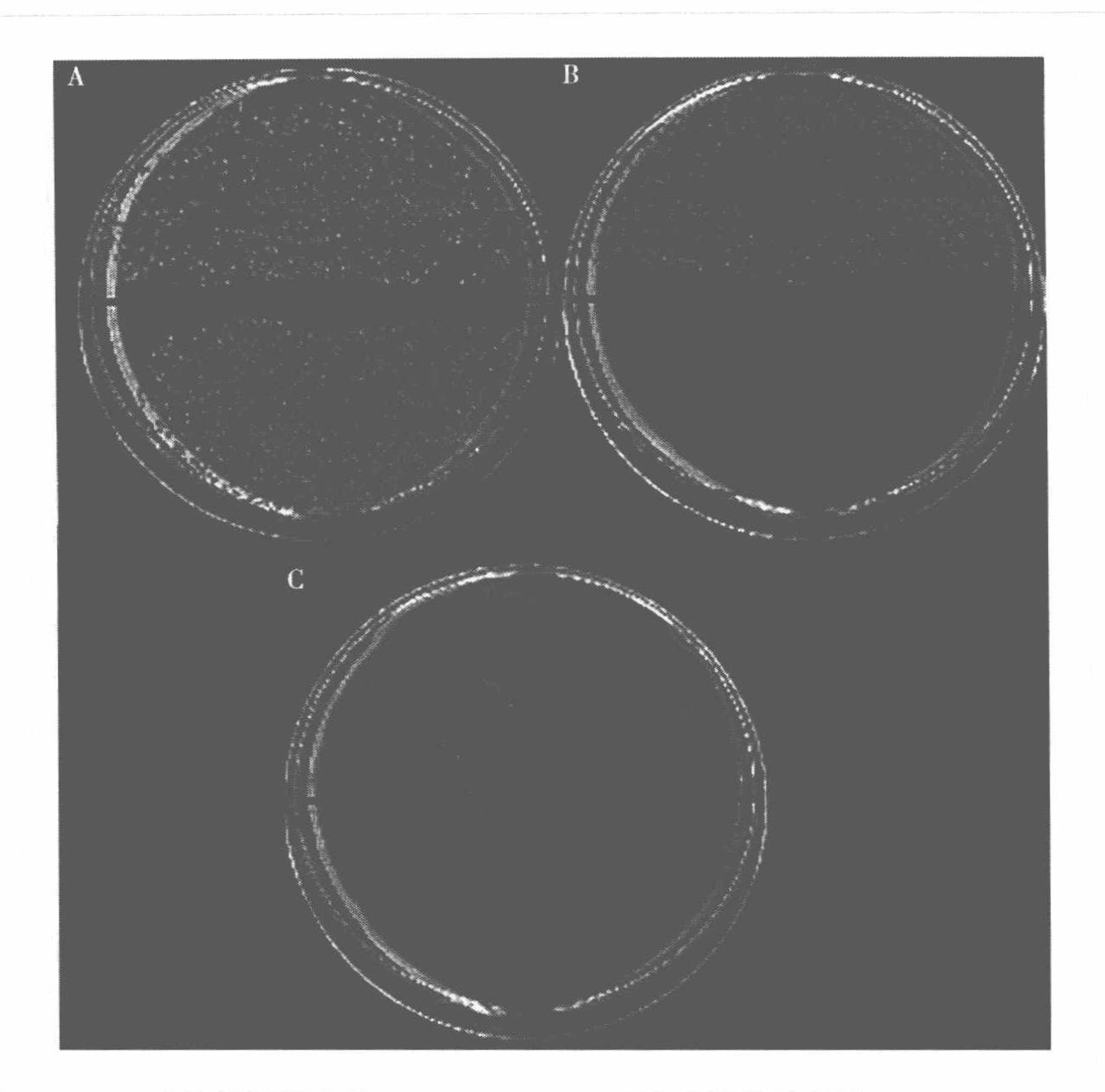


图 15.2　由半胱氨酸蛋白酶（caspase）激活的脱氧核糖核酸酶（CAD）的结构域 *cad* 以及抑制结构域 *icad* 被分别融合到 F［1，2］和 F［3］，F［3］Ile114Val 或 Ile114Ala 中。这些 *cad* 的构建被共同转化到含有 pQE32Δ*icad*-F［1，2］或 *raf*-F［1，2］RBD 的 BL21pREP4 中，并分别在含有选择培养基的培养皿的上部或是下部铺板，在 30℃ 培养 24h。A. pQE-32 *cad*-F［3］I114（野生型 DHFR 片段 F［3］）；B. pQE-32 *cad*-F［3］I114V；C. pQE-32*cad*-F［3］I114A。如副标题 15.3.1.2 中所述的，特别是在对照 1 和 2 中所描述的，共同转化的融合蛋白 *cad* 和 *raf* 的 RAD 起到了内部参照的作用。这些测试使得检测 PCA 的灵敏度限制成为可能。因为在 *cad* 和 *raf* 的 RAD 之间不存在显著地相互作用，PCA 也就不会检测到相互作用。因此，在选择性培养基上也就不会观察到菌落。由于 Ile114Val 变异的引入使得共同转化了相关的 cad 和 icad 构建的菌株可以生长，并且不同于 A 中所述的野生型 F［3］构建，Ile114Val 不会导致共同转化的 *cad* 和 *raf* 的 RAD 的菌落的生成，因此，Ile114Val 变异的运用是理想的选择。F［3］Ile114Ala 变异株不造成阳性对照 *icad*-*cad* 或阴性对照 *icad*-*RBD* 融合蛋白对的生长。

拉链系统。这种灵敏度的变化可以通过一步筛选的选择因子（selection factor）来衡量，选择因子等于共转化细胞数除以在选择压力下存活菌落数。数值越高严密性越高，所以在合理数量的重复竞争下，可以挑选出最好的相互作用亮氨酸拉链异源二聚体[32]。

（3）片段置换（fragment swapping）：无论将相互作用的两个蛋白质分别与 PCA 系统的哪个片段融合，理论上应不影响两个目的蛋白的相互作用。因此，置换与两个目的蛋白质融合的片段应得到类似可比的结果。

（4）无相互作用蛋白质（noninteracting protein）：如果已知一个蛋白质不和任何

一个用于 PCA 测试的目的蛋白有相互作用，理应检测不到 PCA 反应（图 15.2A），并且单独过量表达这个蛋白质也不会对已知的相互作用有竞争影响。

（5）通过竞争实验滴定和降低报告蛋白的酶活：报告蛋白的活性应随着两个融合蛋白表达比例的变化而改变；而且同时单独过量表达相互作用蛋白的其中之一应减弱 PCA 反应的活性。然而，应谨记每个融合蛋白相应的可溶性和稳定性、相互作用蛋白质间的亲和力对比它们的胞内浓度以及 PCA 的灵敏度都会对 titrate 报告蛋白的活性造成影响。因而可以通过降低融合蛋白的表达水平或整合对照 2 和对照 6 来调节报告蛋白的活性，从而降低互补效率。

（6）破坏相互作用：可以预测，在形成复合体的一个单体中插入特定的点突变或删除突变破坏或减弱目的蛋白相互作用力也会影响 PCA 反应。

如果蛋白质工程项目中用于研究的模型性质已经相当清楚，那么只有对照 1、对照 2、对照 4 和对照 6 对于确定实验的特异性和准确性是必需的。RBD-*ras* 的互补突变实验及突变对亲和常数 K_d 的影响已有相关报道，基于这些数据我们构建了若干突变，使 RBD-*ras* 相互作用的 K_d 值降低了三个数量级（图 15.3 例子）。我们将这些突变体和其他突变体都用于 DHFR-PCA 测试，以确保这项实验可以在 1μmol/L 的数量级上检测到 RBD 与 *ras* 的相互作用。另外，已发表的降低蛋白质稳定性的突变，如将核心区疏水氨基酸残基（Val、Leu 或 Ile）突变成 Ala，可以用于严密性测试。

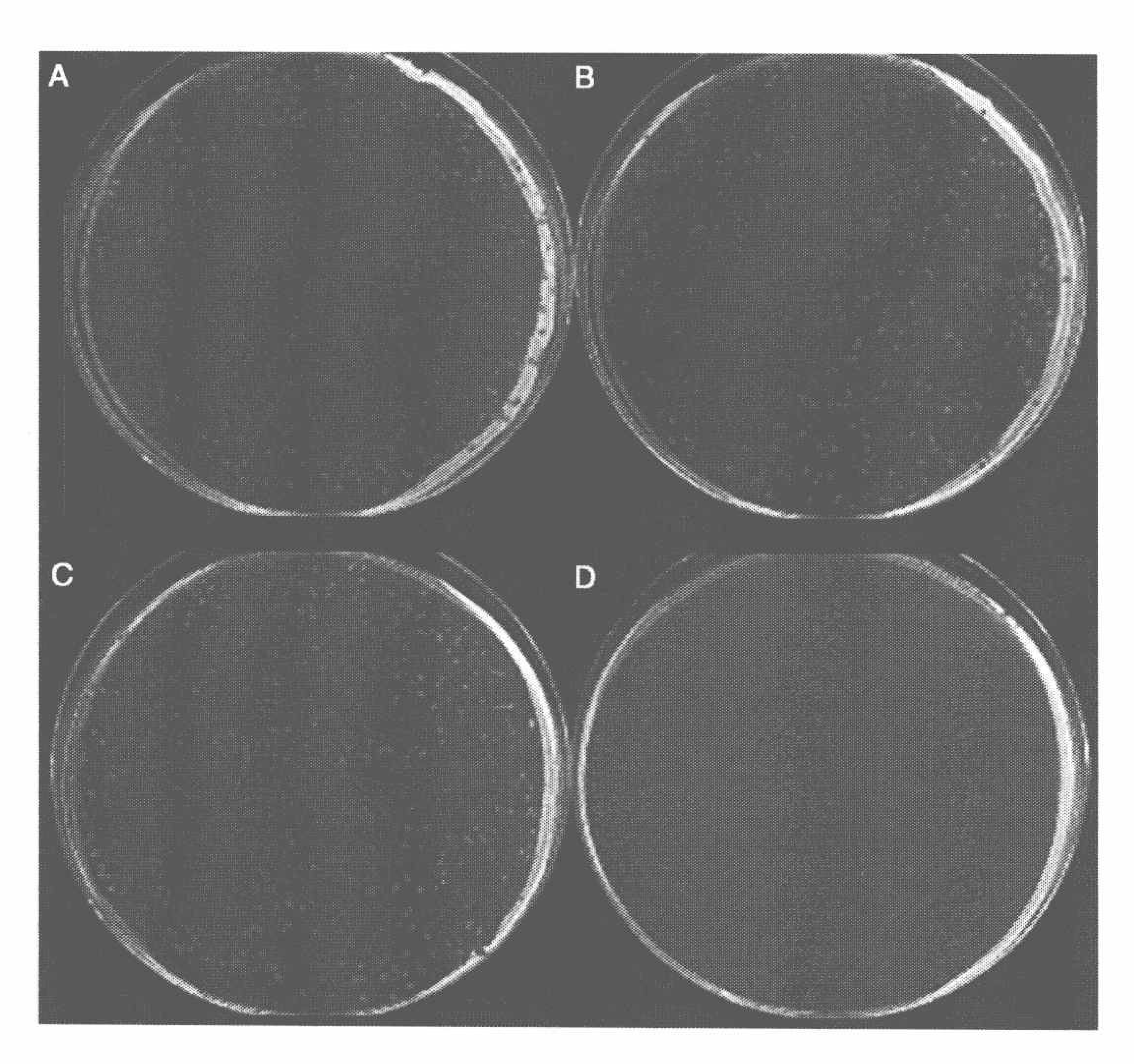


图 15.3 将表达 ras-F［3］的质粒同野生型 raf 的 RBD 或是变异型的 F［1，2］一起转化到 BL21pREP4 菌株中。细菌在选择性培养基中生长 48h，然后培养皿用来筛选。A. 野生型的 *raf* 的 RBD（K_d=0.13μmol/L）；B. K65M 变异的 *raf* 的 RBD（K_d=0.40μmol/L）；C. V69A 变异的 *raf* 的 RBD（K_d=0.95μmol/L）；D. L89L 变异的 *raf* 的 RBD（K_d>0.40μmol/L）。这些变异以及其他一些没有展示的变异使我们能够在 10～100μmol/L 范围内进行试验检测。

15.3.2　合成文库

(1) 为了得到一个无偏倚的文库，首先要构建一个模板，用终止密码子替换基因插入区（可变区），同时插入移码（frame shift）和特定的限制性内切核酸酶位点以确保其明确鉴定（见注 1）。

(2) 要得到每个文库，需要两个部分重叠（一般 18～20bp）的 PCR 产物。例如，对于 PCR 反应 1，使用的一条引物与载体的启动子区域（起始密码子上游 120bp，图 15.4）杂交互补，另一条引物与简并靶标的 5′段互补。对于 PCR 反应 2，需要一条双臂引物和一条与 F［1，2］（可读框 3′下游 120bp 处，图 15.4）杂交的引物。通常，PCR 反应程序设置如下：94℃热起 1min；然后，94℃ 30s，52℃ 30s，72℃ 30s（见注 2），共 25 个循环；最后，72℃延伸 10min 以确保延伸完全。

(3) PCR 产物进行琼脂糖凝胶电泳分析，使用 Gelstar™ 和 Dark reader™（见注 3）观察琼脂糖凝胶上的 PCR 产物，这里使用的是 400～500nm 的蓝光，此波长对 DNA 没有损伤。如果得到目的条带，将剩余的 PCR 产物用于切胶回收。

(4) 使用 Qiaex™ II 纯化切胶产物（见注 4）。

(5) 将来源于 PCR 反应 1 和 2 的产物约 300ng 混合（图 15.4 和注 5），加入 0.2μmol/L 的末端引物（分别与启动子区和 F［1，2］互补）分别与 PCR 反应 1 和 2 的 5′端和 3′端退火。PCR 反应 3 程序如下：94℃热起 1min；然后，94℃ 20s，52℃ 30s，72℃ 30s，共 10 个循环；最后，72℃延伸 10min 以确保延伸完全（见注 6）。

(6) 使用合适的限制性内切核酸酶（此例是 *Sph*I 和 *Xho*I）酶切以上得到的 PCR 产物和载体 pQE-32ΔF［1，2］（见注 7）。

(7) 纯化酶切产物同第（4）步。

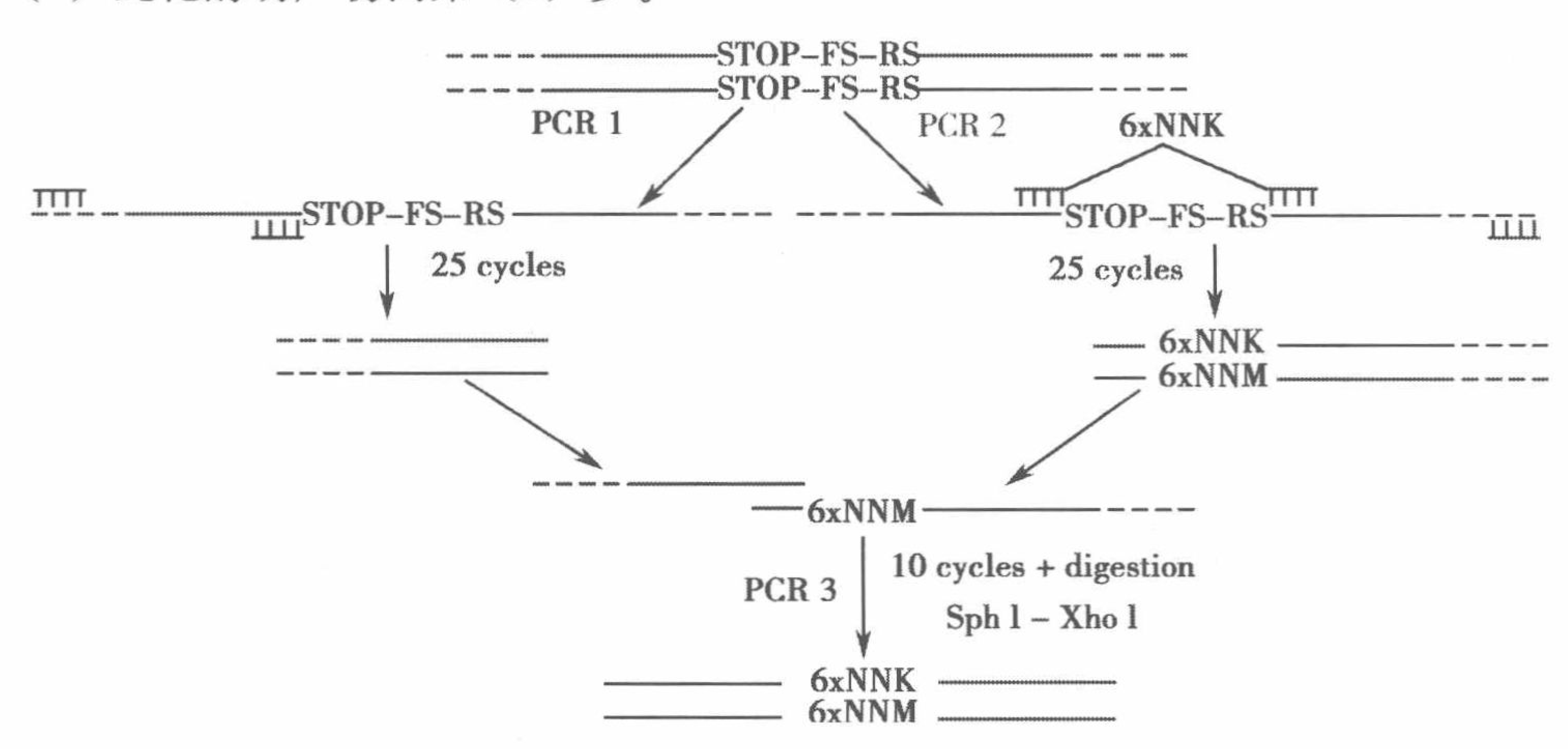


图 15.4　简并文库合成策略的示意图（详见副标题 15.3.2）。合成策略可以分为三个独立步骤：首先，要构建一个可以识别变异的模板（图中实线和虚线分别代表感兴趣的基因和载体）。这个模板是通过用终止密码子、移码（FS）和特定的限制性内切核酸酶位点（RS）替换可变区域的野生型序列来构建。接下来，两个 PCR 反应以这个模板为基础得以执行。PCR1 的产物对应于基因的 5′端，而 PCR2 的产物对应于基因的 3′端。在后一种情况中，模板中的 STOP-FS-RS 序列被适当数量的简并密码子 NNK 所取代。值得注意的是第一轮 PCR 的产物 PCR1 的 3′端可与另一产物 PCR2 的 5′端杂交。这样，在 PCR3 中，由第一轮 PCR 来的产物可以互为模板和引物来进行 PCR 反应。在相对较少的循环数之后（通常为 10），全场的简并基因便可通过限制性内切核酸酶 *Sph*I 和 *Xho*I 酶切回收。这样，DNA 文库便可以用于下一步克隆。

15.3.3 文库的克隆和回收

(1) 插入片段和载体的理想比例是 (2∶1) ～ (3∶1)。我们将用于连接的 DNA 浓度限制在 10ng/μl，使用 1mmol/L 的 ATP，16℃连接过夜（见注 8)。

(2) 连接反应体系于 65℃处理，使酶失活，用氯仿抽提，然后用乙醇沉淀，得到的 DNA 放置几分钟风干，最后用 30μl 去离子水重悬。

(3) 同一天还要制作 SS320 大肠杆菌细胞株（见注 9）的电转感受态（见 15.3.8)。第 (2) 步得到的连接产物与 300μl SS320 细胞混合，加入至槽宽 2mm 的电转杯，使用 Genepulser™ II 进行电转化。将仪器参数设定如下：2.5kV，25μF，200～400Ω。为了得到最好的结果，时间常数对应 200Ω 时应设为 3.8～4.5，400Ω 时应设为 7.6～9.0。脉冲后立即加入 1ml 预冷的 SOC 培养基。将细胞转移至 15ml 锥形管，用 SOC 培养基洗涤电转杯 2 次，最大限度地回收电转细胞，细胞在 37℃适当转速下于 5ml SOC 培养基 [见 15.2.2 (7)] 中复苏 30min。

(4) 取适量电转细胞，稀释 10^4 倍，涂平板，菌落计数，用以衡量连接和克隆的效率（见注 10)。剩余细胞加入至 250ml 的含合适添加剂的 LB 培养基中 [见 15.2.2 (9)] 培养用于提取 DNA。

(5) 使用 Qiagen Midiprep Kit 或其他类似碱裂解试剂盒提取 DNA[47]。

15.3.4 文库筛选

(1) 使用槽宽 1mm 的电转杯，将 15.3.3 (5) 得到的文库 100ng 电转化到 65μl 已经携带有质粒 pQE-32 *ras*-F [3] 的 BL21 pREP4 细胞中。仪器参数：1.25～1.6kV，25μF，200Ω。时间参数对应于 Genepulser™ II 时设为 3.7～4.2，对应于 Electroporator 2510 时设为 4.0～4.6。细胞在 37℃ SOC 培养基中复苏 30min。

(2) 用 PBS 洗细胞 2 次以除去 SOC 培养基。

(3) 将细胞涂布于 15.2.3 (6) 所述的选择性平板，30℃培养 24～72h（见注 12)。同时，另取出部分电转细胞，稀释 10^3 倍涂平板，用于计数和与阳性对照比较，估算转化效率和文库克隆的细胞存活率。例如，我们同时转化相同量的表达野生型 RBD 与 F [1，2] 融合蛋白的载体，将电转细胞稀释 10^4～10^5 倍涂平板。在每一步的操作中，所有的测量都要避免文库与野生型正对照的交叉污染。

15.3.5 克隆竞争实验

改编自参考文献 [32]。

(1) 经过适当时间的培养，用少量的选择培养基回收 15.3.4 (3) 平板中的细胞，加入到 25ml 选择培养基中，30℃ 250r/min 培养。

(2) 培养 24h 后，取 1μl 饱和培养物稀释至 2ml 新鲜选择培养基中。

(3) 可以重复第 (2) 步直至文库达到目标丰度。正常情况下，12 代（12 D）后，文库在很大程度上得以浓缩。但是，这对于不同的系统是不同的，受到各种因素的影响，如文库的简并性和严密性突变体的使用（见 15.3.1.2)。

(4) 在任何一步，都可以将饱和培养物稀释 10^4～10^5 倍涂于选择性平板上，以定性检测竞争效率。随着连续竞争的进行，菌落间的大小差异将逐步减小，菌落的平均

大小将增加。

（5）在任何一代，都可以取出 10μl 文库加入 2ml LB（100μg/ml 氨苄青霉素，25μg/ml 卡那霉素）培养过夜，然后使用 QIAprep™ 提取文库中的克隆的 DNA（见注 14）。

15.3.6　克隆的分离与测序

以下步骤既适用于提取文库中所含的独立克隆的质粒，也适用于上述操作中质粒混合物的提取。

（1）从克隆库中取出 300ng DNA，加入限制性内切核酸酶消化，所用限制性内切核酸酶位点在 pREP4 和 pQE-32 *ras*-F［3］质粒中存在，但在文库载体中不存在。基于此目的，我们使用 *Hpa*Ⅰ、*Xma*Ⅰ、*Eco*NⅠ和 *Xba*Ⅰ（见注 15）。

（2）取 1/10 的酶切产物，转化 XL-1 Blue 感受态细胞，取 20μl 涂布含 100μg/ml 氨苄青霉素的 LB 平板（见注 15）。

（3）挑菌落于含 100μg/ml 氨苄青霉素的 LB 中培养。

（4）提取高质量的 DNA 用于测序。若操作 96 个样品，使用 Montage™ kit；样品数量少，适用 QIAprep™ column kit。

（5）我们使用只能和文库质粒退火（如 F［1，2］内部序列）的引物进行测序（见注 16）。

（6）测序。

15.3.7　蛋白质提纯及其性质检测

分析测序结果，选择感兴趣的克隆排列在 96 孔板的合适位置。此时，克隆可以转化 XL-1 Blue 细胞，冻存作为备份。

（1）选中的克隆重新构建，只与 6×His tag 融合并表达蛋白质（见注 17）。诱导检测 6×His 克隆的表达（见注 18）。

（2）提取正确表达的克隆的质粒。

（3）转化 BL21 pREP4 细胞，涂布含 100μg/ml 氨苄青霉素和 25μg/ml 卡那霉素的平板，37℃培养过夜。若平行操作多个克隆，可以使用 24 孔板。本例中，每个转化所用感受态细胞不要超过 20μl，每平板最多也只涂 20μl。涂板若超过最大允许体积，细胞将不能完全被培养基吸收。

（4）次日，挑菌落于 2.5ml 含合适抗生素的 LB 中 37℃培养过夜（见注 19）。

（5）以 1∶10 的比例将饱和培养物加入 25ml 含合适抗生素的 TB 培养基中（见注 20）。

（6）加入 IPTG 至终浓度为 1mmol/L，37℃培养 90～120min，收获细胞直接提取蛋白质或－80℃保存（见注 21）。

（7）细胞用 1ml Buffer A［见 15.2.6（8）］重悬，然后重新排列在 96 孔板中（见注 22）。3200*g* 离心 40min，去除大部分不可溶组分。

（8）0.2μm 96 孔 PVDF 板中含有 200μl 50%的 Ni-NTA Superflow™ 树脂。将体积扩增器安装于 0.25mm 的 96 孔玻璃纤维过滤板上方，再将此装置安装在真空歧管的密封块上（见注 23）。然后，在第一个过滤板中加入 900μl 样品，再加入 100μl 乙醇以减小交叉感染的风险。用大约 500mbar① 的压力抽滤直至所有样品全部通过此板（见注 24）。

① 1bar＝0.1MPa。

(9) 现在PVDF板应含有树脂和样品抽提物。停止抽滤，从收集器上取下PVDF板，安装在真空歧管的密封块上，然后用大约100mbar的压力抽滤直至所有样品全部通过树脂（见注25）。

(10) 用800μl Buffer B洗2次［见15.2.6（9）］，每次真空压力设为500mbar（见注26）。

(11) 放置好收集器，在100mbar的压力下用100μl Buffer E洗脱样品4次［见15.2.6（10）和注26］。加入1mmol/L DTT并将pH调至5（见注27）。至此，样品可以直接用于性质检测（见注28）。

15.3.8 制备SS320和BL21 pREP4 pQE-32 ras-F［3］感受态细胞

按参考文献［48］方法准备细胞。

(1) 挑SS320或BL21 pREP4 pQE-32 ras-F［3］单克隆于5ml LB（BL21）或5ml SOB（SS320）培养基中，37℃培养5h到过夜。

(2) 取2.5ml培养物加入至500ml LB或500ml SOB培养基中，37℃300r/min培养到OD_{600}为0.5～0.7。

(3) 冰浴10～15min然后转移至预冷的1L离心瓶中。

(4) 5000*g*离心20min。

(5) 弃上清液，用5ml冰水重悬菌体，加入500ml冰水混匀，离心同第（4）步。

(6) 立即弃上清液，用剩余液体重悬菌体。

(7) 再加入500ml冰水混匀，离心同第（4）步。

(8) 立即弃上清液，用剩余液体重悬菌体。

(9) 若细胞用于电转化，将重悬细胞转移至预冷的50ml离心管，2℃ 5000*g*离心10min。估算菌体体积（一般500ml培养物会得到500μl），在冰上加入相同体积的预冷水重悬细胞。Aliquot 50～300μl细胞至预冷的离心管中，细胞密度大约是2×10^{11}个细胞/ml。

(10) 若要冻存电转化细胞，在第（8）步得到的细胞中加入40ml预冷的10%甘油并混匀。离心细胞同第（9）步，估算菌体体积，在冰上加入相同体积的预冷10%甘油重悬细胞。Aliquot 50～300μl细胞至预冷的离心管中用干冰冷冻，－80℃储存。

15.3.9 制备XL-1 Blue和BL21 pREP4感受态细胞

按照参考文献［49］方法准备细胞，稍作改动如下：细胞在第一次离心后仅洗1次，4℃倒置离心瓶于纸上，除去痕量的培养基（见注29）。

(1) 在200ml SOB或LB培养基中加入过夜培养物，18℃（200～250r/min）培养至OD_{600}约0.6。

(2) 冰浴10min并转移至500ml离心瓶。

(3) 4℃ 2500*g*离心1min。

(4) 用80ml预冷转化缓冲液重悬菌体，并冰浴10min，离心同（3）。

(5) 用20ml预冷转化缓冲液轻柔的重悬，加入DMSO至终浓度为7%。

(6) 冰浴10min。

（7）Aliquot 细胞重悬物，液氮冷冻。

15.4 注

1. 在 PCR 之前，用 T4 多核苷酸激酶（polynucleotide kinase）磷酸化引物使其可以进行连接反应。两个引物分别与被删区的两边配对。使用高保真的聚合酶，如 Pfu 或 Pfu Turbo，进行 10～16 个循环的 PCR 反应。所得的 PCR 产物是不包含被删区的线性质粒。使用 *Dpn*I 消化 PCR 产物，然后用乙醇沉淀，取 1/10 的 PCR 产物用于连接和转化，并筛选阳性克隆。插入移码是为了避免终止密码子被读通。PCR 方法来源于 ExSiteTM PCR 定向突变试剂盒（Stratagene）。

2. 因为要扩增的序列很短，我们使用 *Taq* 聚合酶。若要扩增长基因，推荐使用高保真的聚合酶。如果 PCR 反应 1 与 PCR 反应 2 产物的大小差异很大，特别是对于大的基因，推荐使用 mega-primer 方法[50,51]。这两个方法稍作改变，都可以用于构建序列中相距较远的区域同时变化的复合文库。

3. 将 Gelstar™ 稀释 10^4～10^5 倍用于 TAE 琼脂糖凝胶电泳，Gelstar 比 EB 更易分解，所以使用 Gelstar 须当天制胶。而且，Gelstar™ 信号在 DNA 浓度较低的条件下即饱和，所以高于一个浓度就很难定量 DNA 了。另外，Gelstar™ 对样品迁移速度也有较大影响，所以要注意每孔的上样量（一般来说，25μl 的上样孔不要超过 300ng 质粒 DNA）。如果用 EB 染色也可以在蓝光下观察，但是若片段较小或是 PCR 产物则需要的样品量较多（线性质粒 DNA 要大于 1.5μg）。当然也可以在紫外光下进行观察，但是要格外小心。

4. 同时构建几个文库时，更倾向于使用 QIAquick™ 凝胶回收的方法，但是比较贵。也可以用 *Dpn*I 37℃酶切 PCR 产物 2h 以去除模板质粒 DNA，从而省去了 15.3.2 中第（3）步的电泳切胶的步骤，然后 PCR 产物用 QIAquick™ 方法回收。

5. 这个量可以根据所研究基因的大小成比例增加。加入到 PCR 反应 3 中的 PCR 反应 1 和 2 产物的相对量可以根据它们的分子质量大小进行适当调节。

6. PCR 反应 3 的循环数要降到最小，否则会影响文库的重复性（图 15.3）。这一步值得优化以在最小循环数的条件下得到最大量的产物，我们推荐使用注 2 中提到的 *Taq* 聚合酶。

7. 这个载体是使用 *Nde*I 和 *Xba*I 消化 pQE-32 然后重新连接得到的，去除了终止子与复制子起点之间约 850bp 的片段。根据我们的实验结果，蛋白质在这个新载体中的表达量和稳定性与之前一样，但是新载体的转化效率提高了一个数量级。而且，去除了原载体中 *Xho*I 位点，这样构建文库时就可以使用这个酶了。

8. 当使用的载体没有进行磷酸化修饰时，在此比例下得到的结果最好。

9. SS320 细胞是 MC1061 和 XL-1 Blue 细胞杂交得到的[30]。将等体积 OD_{600} 均为 0.6 的两种细胞混合，37℃低转速（50r/min）培养 1h，然后将转速升至 250r/min 培养 5min，终止结合。涂布添加了 10μg/ml 四环素和壮观霉素的 LB 平板，分离新菌株。这个菌株不仅具有 MC1061 的高转化效率（大于 5×10^{10} 菌落/μg 超螺旋质粒），并且也具有来源于 XL-1 Blue 的转化 pQE 载体所需的 $LacI_q$ 过量表达的性质。也可以用其他适用于给定表达系统的菌株。

10. 菌落计数并乘以稀释倍数。用 30μl 水重悬 300ng 的质粒，转化入 300μl 的

SS320 感受态细胞中，一般可以得到 10^6～10^7 个菌落。

11. 我们最初选择 pQE 载体是因为我们希望可以在进行筛选和 Ni-NTA 亲和纯化过程之间无需更换载体。使用 pQE-32 衍生载体表达 ras-F [3] 融合蛋白[31]，虽然该质粒具有和 *raf* 的 RBD 文库相同的复制起点和抗生素选择标记，但是在三甲氧苄二氨嘧啶的选择压力下两个质粒都不会丢失（如 mDHFR 的重构需要两个融合蛋白同时表达）。实际上我们构建了一个起始位点和抗性标记不同的 pQE-32 衍生质粒用来表达 *ras* 融合蛋白，但是使用该质粒大大降低了筛选的严密性，在这种条件下，基于序列信息，DHFR PCA 选出的克隆没有结合 *ras* 的能力。我们得到这个结论是因为，DHFR PCA 筛选所得克隆与 *ras* 结合的重要区域的序列中含有终止密码子和异常序列，说明这些相互作用是不特异的。这个问题可能可以通过在 F [3] 中插入不稳定突变 I114V 或 I114A 得以解决，但我们还没有检验。我们认为我们选择的方法严密性较高，因为在进行文库转化之前携带 pQE-32 *ras*-F [3] 的 BL21 pREP4 细胞有着 ColE1 复制起点的最高拷贝数，使携带的文库质粒拷贝数最小，这对细胞在选择培养基上生长是必要的。将来，可以使用低拷贝的表达诱饵蛋白的载体或 DHFR 不稳定突变[31,32]。

12. 也可以直接将电转细胞加入 25ml 的 M9 培养基中，这对直接开始竞争实验特别有用。

13. 观察发现，竞争实验前将细胞涂平板可以改善严密性，特别是在第一代，这意味着在固体培养基上生长可以加快最高克隆的选择效率。

14. 这个阶段我们推荐使用 QIAprep™ spin kit，因为用碱裂解法在 BL21 pREP4 中提的质粒质量不够好。另外，提质粒时用 Qiagen PB Buffer 洗 1 次柱子。

15. 第 2、第 3 步免去了测序前检查细胞中是否含有正确质粒的必要。本例中按此程序处理的 700 个菌落中含有 pQE-32 *ras*-F [3] 或 pREP4 质粒。

16. PCR 反映 2 和 3 中所用的 3′引物相同 [见 15.3.2 (5)]。

17. 建议体外检测蛋白质性质之前去除 DHFR 片段，除非特定的实验需要，因为它会减小产量，使提纯步骤更复杂，也有可能改变蛋白质性质。建议将蛋白质重新构建到不含 DHFR 片段的质粒中。为了实现这一目标，我们选用了兼容的酶切位点，用 *Xho*I 和 *Sal*I 分别克隆文库和 F 片段 DNA [1, 2]。

18. 过夜饱和培养物 1∶10 稀释至含 100μg/ml 氨苄青霉素的 TB 培养基中，37℃ 300r/min 培养 90min，用 1mmol/L 的 IPTG 诱导 4h。然后每个克隆取出 60μl，加入等体积 2×SDS-PAGE 上样缓冲液，电泳分析。可用考马斯亮蓝染色，或用抗 Histag 的抗体蛋白质印迹法检测蛋白质表达情况。

19. 在大规模提纯之前，建议测试几个克隆不同体积的产量。我们建议在变性条件下提纯蛋白质，是因为同时进行多个平行提纯时，它更易于操作。但是，如果蛋白质可溶性很好，也可以在非变性条件下提纯。本例中，表达 *raf* RBD 突变体的 25ml 的 BL21 pREP4 细胞培养物，按 15.3.7 所述处理，可以得到 400μl 浓度 200～600μg/ml 的蛋白质。

20. 根据所需蛋白质的量或蛋白质表达水平，培养物体积可为 5～50ml，详见注 19。

21. 蛋白质也可以 30℃诱导表达过夜。

22. 除此之外，若用于分析的克隆量少，可以使用 Ni-NTA 旋转柱。如果需要独立克隆的较多蛋白质，可选择 Ni-NTA 琼脂糖柱纯化。本例中，在非变性条件下纯化是有用并合理的。

23. 所述系统仅适用于 Millipore 的大收集器和密封模块。离心前可以抽真空。

24. Millipore Manifold 中包含一个测量仪表 gauge。

25. 若需要，收集各漂洗步骤的穿出液和滤过液［见 15.3.7（9）和（10）］。

26. 本例中，蛋白质在变性条件下被洗脱，然后将样品稀释至变性剂浓度不会对蛋白质功能造成影响。需根据所研究蛋白质的稳定性来确定洗脱缓冲液中的变性剂浓度。但是，若要求在天然状态下洗脱蛋白质，Qiagen 推荐使用浓度梯度递减的盐酸胍。这样用合适的缓冲液洗脱出的蛋白质中就不含变性剂了。当然也可以在非变性条件下提纯蛋白质（见 Qiagen Expressionist and Ni-NTA spin column handbook）。我们使用 25mmol/L 或 50mmol/L 的乙酸缓冲液（pH5.0）改善洗脱效率（见注 27，洗脱后适当调节 pH）。本研究中，合适浓度的尿素也可以用作变性剂。另外，因为尿素在用氰酸盐平衡的溶液中，使用时应小心。但是，高浓度的尿素也不会沉淀 SDS，所以每一步纯化所得的样品都可以用于 SDS-PAGE 分析。

27. 洗脱后在样品中加入 DDT 是因为 Ni-NTA 不推荐使用 DDT。可以用 NaOH 或其他合适弱碱，如 Tris 或 NaOAc，调节洗出液 pH。

28. 如果样品不是立即用于性质鉴定，我们建议在样品中加入 40%（*V*/*V*）的甘油和 1mmol/L NaN_3，－20℃保存。在性质鉴定前通过透析或超滤去除甘油。或者，不加甘油，用液氮速冻样品，－80℃保存。

29. 根据 QIAEXpressionnist Handbook，用 TFB 缓冲液（RbCl）制作的 BL21 pREP4 感受态细胞效价更高，但是由于我们仅用该细胞转化超螺旋 DNA，用 Inoue 方法制作感受态细胞即可[48]。

致谢　感谢 Emil Pai、David Waugh 和 Shigekazu Nagata 分别对 h-*ras*、*raf*RBD 以及 CAD 的 *cad* 和 *icad* 结构域的 cDNA 给予的帮助。并感谢 Dimitri Sans 对本书手稿的仔细阅读，Jérôme Dupras 对 CAD 的构建提供的帮助，以及 Joelle Pelletier 和实验室其他成员对开发 DHFR PCA 技术作出的贡献。F.-X.-C.-V. 是加拿大健康研究所研究人员培训基金。

参考文献

1. Chen, K. Q. and Arnold, E H. (1991) Enzyme engineering for nonaqueous sol-vents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Biotechnology* (*NY*) 9, 1073-1077.

2. Chen, K. Q., Robinson, A. C., Van Dam, M. E., Martinez, P., Economou, C., and Arnold, F. H. (1991) Enzyme engineering for nonaqueous solvents. II. Additive effects of mutations on the stability and activity of subtilisin E in polar organic media. *Biotechnol. Prog.* 7, 125-129.

3. Iffland, A., Tafelmeyer, P., Saudan, C., and Johnsson, K. (2000) Directed molecular evolution of cytochrome c peroxidase. *Biochemistry* 39, 10790-10798.

4. Scott, J. K. and Smith, G. P. (1990) Searching for peptide ligands with an epitope library. *Science* 249, 386-390.

5. Lowman, H. B., Bass, S. H., Simpson, N., and Wells, J. A. (1991) Selecting highaffinity binding proteins by monovalent phage display. *Biochemistry* 30, 10832-10838.

6. Keefe, A. D. and Szostak, J. W. (2001) Functional proteins from a randomsequence library. *Nature* 410, 715-718.

7. Imanaka, T., Shibazaki, M., and Takagi, M. (1986) A new way of enhancing the thermostability of proteases. *Nature* 324, 695-697.

8. Shih, P. and Kirsch, J. F. (1995) Design and structural analysis of an engineered thermostable chicken lysozyme.

Protein. Sci. 4, 2063-2072.

9. Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., et al. (1997) Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* 4, 805-809.

10. Kim, D. E., Gu, H., and Baker, D. (1998) The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc. Natl. Acad. Sci. USA* 95, 4982-4986.

11. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262, 1680-1685.

12. Stemmer, W. P. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389-391.

13. Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 391, 288-291.

14. Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat. Biotechnol.* 16, 258-261.

15. Ostermeier, M., Shim, J. H., and Benkovic, S. J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* 17, 1205-1209.

16. Lutz, S. and Benkovic, S. J. (2000) Homology-independent protein engineering. *Curr. Opin. Biotechnol.* 11, 319-324.

17. Coco, W. M., Levinson, W. E., Crist, M. J., et al. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* 19, 354-359.

18. Murakami, H., Hohsaka, T., and Sisido, M. (2002) Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. *Nat. Biotechnol.* 20, 76-81.

19. Crameri, A., Dawes, G., Rodriguez, E., Jr., Silver, S., and Stemmer, W. P. (1997) Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nat. Biotechnol.* 15, 436-438.

20. Christians, F. C., Scapozza, L., Crameri, A., Folkers, G., and Stemmer, W. P. (1999) Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat. Biotechnol.* 17, 259-264.

21. MacBeath, G., Kast, P., and Hilvert, D. (1998) Redesigning enzyme topology by directed evolution. *Science* 279, 1958-1961.

22. Ostermeier, M., Nixon, A. E., Shim, J. H., and Benkovic, S. J. (1999) Combinatorial protein engineering by incremental truncation. *Proc. Natl. Acad. Sci. USA* 96, 3562-3567.

23. O'Neil, K. T. and Hoess, R. H. (1995) Phage display: protein engineering by directed evolution. *Curr. Opin. Struct. Biol.* 5, 443-449.

24. Roberts, R. W. and Szostak, J. W. (1997) RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA* 94, 12297-12302.

25. Firestine, S. M., Salinas, F., Nixon, A. E., Baker, S. J., and Benkovic, S. J. (2000) Using an AraC-based three-hybrid system to detect biocatalysts *in vivo*. *Nat. Biotechnol.* 18, 544-547.

26. Hanes, J., Jermutus, L., and Plückthun, A. (2000) Selecting and evolving functional proteins *in vitro* by ribosome display. *Methods Enzymol.* 328, 404-430.

27. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that dis- play cloned antigens on the virion surface. *Science* 228, 1315-1317.

28. Dunn, I. S. (1996) Phage display of proteins. *Curr. Opin. Biotechnol.* 7, 547-553.

29. Forrer, P., Jung, S., and Plückthun, A. (1999) Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. *Curr. Opin. Struct. Biol.* 9, 514-520.

30. Sidhu, S. S., Lowman, H. B., Cunningham, B. C., and Wells, J. A. (2000) Phage display for selection of novel binding peptides. *Methods Enzymol.* 328, 333-363.

31. Pelletier, J. N., Campbell-Valois, F -X., and Michnick, S. W. (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally-design fragments. *Proc. Natl. Acad. Sci. USA* 95, 12141-12146.

32. Pelletier, J. N., Arndt, K. M., Plückthun, A., and Michnick, S. W. (1999) An *in vivo* library-versus-library

selection of optimized protein-protein interactions. *Nat. Biotechnol.* 17, 683-690.

33. Michnick, S. W., Remy, I., Campbell-Valois, F X., Vallee-Belisle, A., and Pelletier, J. N. (2000) Detection of protein-protein interactions by protein fragment complementation strategies. *Methods Enzymol.* 328, 208-230.
34. Michnick, S. W. (2001) Exploring protein interactions by interaction-induced fold- ing of proteins from complementary peptide fragments. *Curr. Opin. Struct. Biol.* 11, 472-477.
35. Appleman, J. R., Prendergast, N., Delcamp, T. J., Freisheim, J. H., and Blakley, R. L. (1988) Kinetics of the formation and isomerization of methotrexate complexes of recombinant human dihydrofolate reductase. *J. Biol. Chem.* 263, 10304-10313.
36. Arndt, K. M., Pelletier, J. N., Muller, K. M., Alber, T., Michnick, S. W., and Plückthun, A. (2000) A heterodimeric coiled-coil peptide pair selected *in vivo* from a designed library-versus-library ensemble. *J. Mol. Biol.* 295, 627-639.
37. Mossner, E., Koch, H., and Plückthun, A. (2001) Fast selection of antibodies without antigen purification: adaptation of the protein fragment complementation assay to select antigen-antibody pairs. *J. Mol. Biol.* 308, 115-122.
38. Michnick, S. W. and Shakhnovich, E. (1998) A strategy for detecting the conservation of folding-nucleus residues in protein superfamilies. *Fold. Des.* 3, 239-251.
39. Campbell-Valois, F X., Tarassov, K., and Michnick, S. W. (2005) Massive sequence perturbation of a small protein. *Proc. Natl. Acad. Sci. USA* 102, 14988-14993.
40. Gu, H., Yi, Q., Bray, S. T., Riddle, D. S., Shiau, A. K., and Baker, D. (1995) A phage display system for studying the sequence determinants of protein folding. *Protein Sci.* 4, 1108-1117.
41. Emerson, S. D., Waugh, D. S., Scheffler, J. E., Tsao, K. L., Prinzo, K. M., and Fry, D. C. (1994) Chemical shift assignments and folding topology of the Ras-binding domain of human Raf-1 as determined by heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* 33, 7745-7752.
42. Emerson, S. D., Madison, V. S., Palermo, R. E., et al. (1995) Solution structure of the Ras-binding domain of c-Raf-1 and identification of its Ras interaction surface. *Biochemistry* 34, 6911-6918.
43. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Bacterial media, antibiotics and bacterial strains, in *Molecular Cloning* (Nolan, C., ed.), vol. 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. A. 3.
44. Remy, I., Wilson, I. A., and Michnick, S. W. (1999) Erythropoietin receptor activation by a ligand-induced conformation change. *Science* 283, 990-993.
45. Nassar, N., Horn, G., Herrmann, C., Scherer, A., McCormick, F., and Wittinghofer, A. (1995) The 2.2 A crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue. *Nature* 375, 554-560.
46. Block, C., Janknecht, R., Herrmann, C., Nassar, N., and Wittinghofer, A. (1996) Quantitative structure-activity analysis correlating Ras/Raf interaction *in vitro* to Raf activation *in vivo*. *Nat. Struct. Biol.* 3, 244-251.
47. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Alkalyne lysis and PEG preparation for large scale DNA preparation, in *Molecular Cloning*, vol. 1 (Nolan, C., ed.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1.38-1.41.
48. Seidman, C. E., Struhl, K., Sheen, J., and Jessen, T. (1997) Introduction of plasmid DNA into cells, basic protocol 2, in *Current Protocols in Molecular Biology*, vol. 1, Suppl 37 (Ausubel, F M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds.) John Wiley & Sons, Inc., New York, NY, pp. 1.8.4-1.8.5.
49. Inoue, H., Nojima, H., and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96, 23-28.
50. Sarkar, G. and Sommer, S. S. (1990) The "megaprimer" method of site-directed mutagenesis. *Biotechniques* 8, 404-407.
51. Brons-Poulsen, J., Petersen, N. E., Horder, M., and Kristiansen, K. (1998) An improved PCR-based method for site directed mutagenesis using megaprimers. *Mol. Cell Probes* 12, 345-348.

16 应用末端截切、进化、再延长技术提高酶稳定性的方法

Jochen Hecky, Jody M. Mason, Katja M. Arndt and Kritian M. Müller

概要 提高酶稳定性是改造酶在高温等极端条件下行使其活性理想的设计步骤，同时也能降低酶对蛋白酶剪切的敏感性。由于这些优点，研究者已经发明了很多不同的方法和技术改造蛋白质，但到目前为止这些方法的效果很有限。这里，我们介绍包括末端截切、随机突变和断裂、重组、再延长，然后在生理温度下筛选，以期获得稳定性提高的酶活突变体。以 TEM-1 β-内酰胺酶为模型蛋白，通过三轮酶的定向进化，包括随机突变、DNA 混编以及在 37℃下的筛选，得到了其体内氨苄抗性与野生型相当的有删截的突变体。动力学研究表明，筛选到的突变体与其野生型相比有显著的热稳定性。这种方法非常有效，进化所产生的酶能够在比野生型的最适温度高 20℃的条件下，保持其最高的酶活反应。因此在生理温度下通过末端截切和接下来作为补偿的定向进化来干扰结构，是提高蛋白质热稳定性的快速、高效的方法，避免了在高温下筛选的需要。
关键词 稳定蛋白质；蛋白稳定性；末端截切；热稳定性；酶活；DNA 混编；随机突变；再延长

16.1 引 言

如果酶能够被改造使得其稳定性得以提高，它将会得到充分的开发利用。例如，酶能够在比较大的温度范围内具有活性，或者对蛋白酶的敏感性降低从而半衰期得到延长。蛋白质稳定性的提高，如同增加其特异性和反应活性一样，在工业和制药业中是最受关注的特性之一。

目前为止，发现了很多提高蛋白质热稳定性的因素[1~4]。例如，刚性和紧密程度的增加，更多的疏水残基在核心中的引入，以及范德华作用力的增加。此外，蛋白质的热稳定性还可以通过引入金属结合位点或额外的二硫键[5,6]、主链环化以及缩短或删除环区[7]等步骤来提高。决定蛋白质高热稳定性的两个主要因素被认为是氢键增加和离子键网络的形成[8~10]。

然而，由于分子水平上对蛋白质热稳定性仍未有完整的认识，把蛋白质改造成热稳定性高的蛋白质仍然是个难题[11]。计算和比较研究法用于识别其稳定作用并已经成功地加以应用[12~14]。然而它们必须依赖于有良好分辨率的晶体或核磁共振结构。结构中能清楚地看到 Cα 的轮廓，并且能获得许多同源序列，最好是嗜热来源的。此外，模拟达尔文优化循环“突变、重组和选择”的进化方法也被用来突破目前蛋白质合理设

计的界限[15,16]。

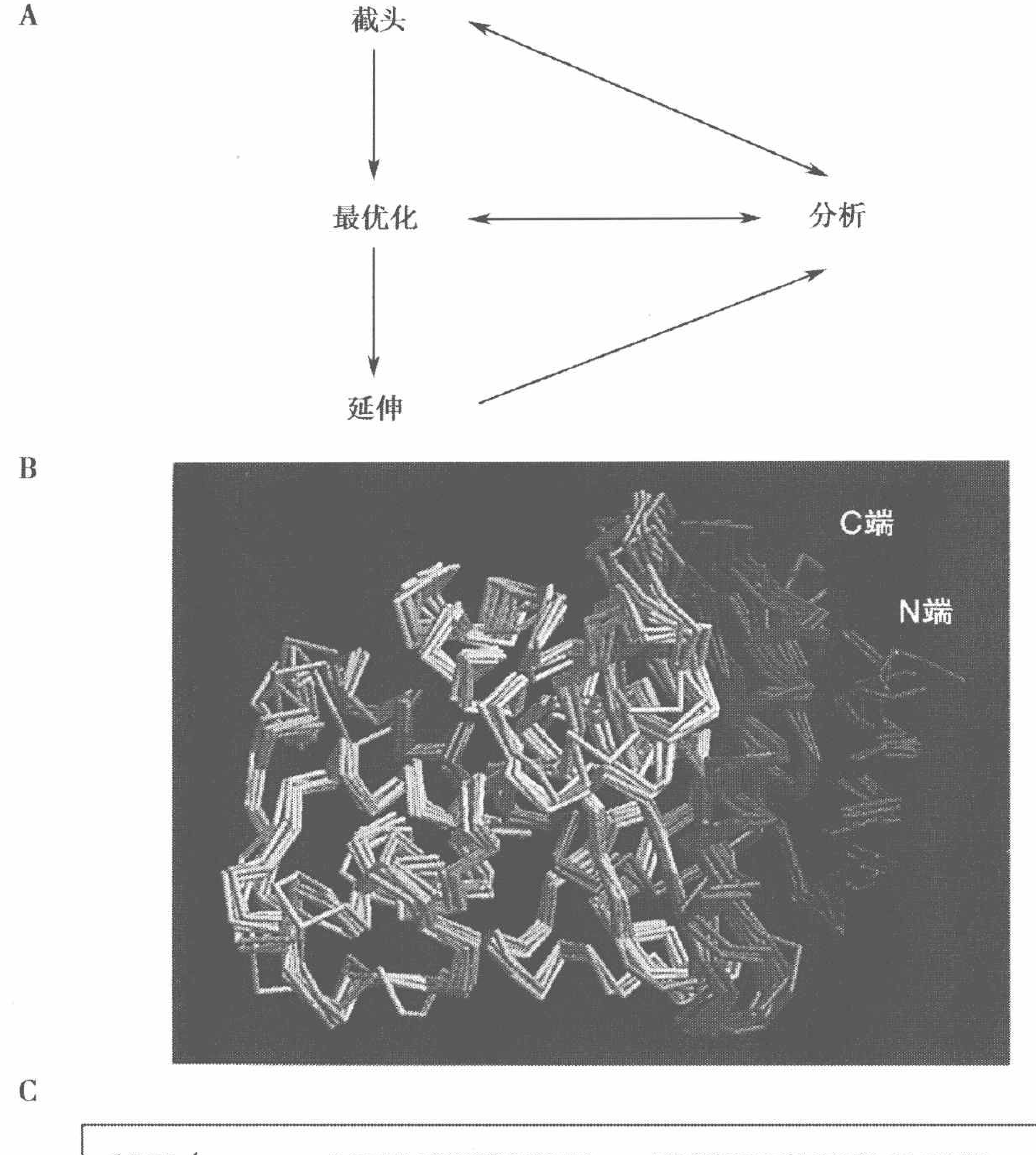
16.1.1　末端截切

由于蛋白质不同，其末端区域对于结构和功能完整性的贡献也有所不同。对有些蛋白质来说，剪切掉末端部分氨基酸并不影响其三维结构和构造稳定性。在这些结构中，其末端部分经常未被定义或是呈无规则卷曲状态。然而，另一些蛋白质对其末端缺失非常敏感。在这种情况下，其末端少数几个氨基酸的缺失就会导致其构象稳定性极大地降低，从而形成松散结构[17]，使其对蛋白酶更加敏感或是更易聚集形成沉淀[18]。这种现象已分别在 RNA 酶 A[19]、RNA 酶 HI[20]、金黄色葡萄球菌核酸酶 R (staphylococcal nuclease R)[17]、硫氰酸酶[21,22]、*Taq* DNA 聚合酶的 Stoffel 片段[23] 以及氯霉素乙酰转移酶[18] 中得到证实。尽管部分蛋白质不能缺失其末端的甚至一个氨基酸，对于绝大多数蛋白质来说，其末端氨基酸的缺失会显著影响其表达水平、原始构象和稳定性，然而在一定范围内仍然有功能（如较低的反应温度）。不过，一旦超越这一范围，末端截切便会导致不可恢复的结构毁坏和功能的完全丧失。因此，仔细选择合适的末端截切可以使绝大多数蛋白质在结构受到干扰的情况下不影响其折叠过程和功能构象。

16.1.2　定向进化互补对结构的干扰

蛋白质结构可以通过替换、插入和删除氨基酸进行干扰。这种干扰可以通过单个或多个通常被称为全局或第二位抑制子[24] 的补救突变来稳定，因为它们能够在远程抑制原本有害突变的表型。这种现象是由于蛋白质结构的高度复杂性和可塑性以及控制蛋白质结构的蛋白质内部相互作用造成的。并且这也是由蛋白质结构的自身简并特性促成的，蛋白质结构往往由少数几个关键氨基酸决定，而绝大多数氨基酸被替换后并不影响其表型变化。被截切 RNase HI 的第二位点抑制因子已通过随机突变引入并且通过功能筛选得到[20]。然而，单纯由随机突变引入的突变，由于其有害突变的积聚而使有功能的蛋白质减少，从而可检测的有益突变将变少。为了克服随机突变的局限性，一个被称为 DNA 混编的重组方法被引进[15,16]。这在基于 DNA 的蛋白质工程中是一个里程碑。它是不断产生并打乱点突变重复循环过程，被称为体外或定向进化。它基于同样支配着自然进化的原则，即随机突变、重组、筛选或选择。用这种方法，很多酶在活性、专一性、稳定性，甚至在有机溶剂中的功能等方面有显著改进[25]。酶的定向进化比起合理设计有很多优点，因为不需要知道三维结构而能提高酶特性。然而，它仅限用于有合适的筛选过程的酶。

本章总结了一个可控制的、有广泛应用前景的提高酶稳定性的方法。它基于在 DNA 水平上对目标蛋白末端进行截切来实施结构干扰（见 16.3.2 和 16.3.3）。由截切导致的有害突变（见 16.3.4）是由进化的、遗传的体外优化过程来克服的（见 16.3.5）。截切-优化-延长的方法（图 16.1A）优越于其他方法之处在于它的简单性，因为它不依赖于结构信息，也不需要知道同源蛋白。然而，如果目前的数据能够提供这样的信息，截切设计的过程将会更直接，在某些情况下甚至更有效。最后，本章所描述的方法还有一个优点，那就是用该方法得到的热稳定突变体能够在生理温度下筛选而得。



C

1BTL/ 1TEM	HPETLVKVKDAEDQL	~	MDERNRQIAEIGASLIKHW
1SHV:A	spQPLEQIKLSESQL	~	MAERNQQIAGIGAALIEHWqr
1MBL:B	*ktemk*DDFAKLEEQF	~	--- DDKLIAEATKVVMKaln*mng*
1BZA	anSVQQQLEALEKSS	~	R--RRDILAAAAKIVThgf
1BUL	ntkGIDEIKNLETDF	~	H--EDKVIAEASRIAIdnlk
1BSG	*ghgsgsvs*DAERRLAGLERAS	~	R--DDGLVADAARVLAETLg
1MFO	-APIDDQLAELERRD	~	N--LRPLIGELTALVLPSLl
1PIO:A	----mKELNDLEKKY	~	P--NDKLISETAKSVMKEF--

图 16.1　A. 关于截切（truncation）-优化（optimization）-延伸（elongation）的图表。酶被截去了头部，接着通过定向演化被优化，最后延伸成全长蛋白质。所有的步骤都可以对它们的活性和稳定性进行分析。B. 表示了来自不同细菌物种的各种各样的β-内酰胺酶（β-lactamases）利用 CE 算法重叠的 Cα 骨架图[30]。在 RasMol 中，相邻的二级结构单元用不同的灰度表示[47]。C. 在 B 中显示的β-内酰胺酶的 N 端和 C 端部分的基于结构的序列比对，与显示的 TEM-1 酶（PDB 号 1btl/1tem）的 N 端和 C 端螺旋相一致。这些序列由 PDB 号表示。由小写字母表示的残基在 CE 算法中是被忽略的，因此没有包含在叠加中。那些在 SWISSPROT 数据库中出现，而在相应的 X 射线结构中没有出现的残基用斜体表示。省略的序列用符号“～”表示。选择的β-内酰胺酶基因显示同源性为 67.8%～32.4%。一致结构的偏差的平均方根值为 1.2～2.2Å（以 TEM-1β-内酰胺酶为参考），并且 Z 得分大于 7。结构的重叠展示了末端螺旋的结构可变性。

16.2　材　　料

16.2.1　质粒构建、DNA 混编和随机突变

(1) Pr _ sfi _ pelB _ DG _ blawt 的正向引物：5′-TTACTCACGGCCCAGCCGGCCATGGCTGACGG**TCACCAGAAACGCTGGTG**-3′，有下画线的是限制性内切核酸酶位点，黑体为杂交区域。

(2) Pr _ sfi _ pelB _ DG _ bladelHPE 的正向引物：5′-TTACTCACGGCCCAGC-CGGCCATGGCTGACGG **TACGCTGGTGAAAGTAAAAAGATG**-3′。

(3) Pr _ sfi _ pelB _ DG _ bladelHPETL 的正向引物：5′-TTACTCACGGC-CCAGCCGGCCATGGCTGACGG **TGTGAAAGTAAAAGATTGCTGAAG**-3′。

(4) Pr _ blawt _ GG _ his5 _ hind 的反向序列：5′-TAGTCAAGCTTACTA GT-GATGGTGATGGTGGCCACC**CCAATGCTTAATCAGTGAGG**-3′。

(5) Pr _ bladelW _ GG _ his5 _ hind 的反向序列：5′-TAGTCAAGCTTACTA GTGATGGTGATGGTGGCCACC**ATGCTTAATCAGTGAGGCACC**-3′。

(6) Pr _ bla _ delKHW _ GG _ his5 _ hind 的反向序列：5′-TAGTCAAGCTTAC-TA GTG ATGGTGATGGTGGCCACC**ATGCTTAATCAGTGAGGCACC**-3′。

(7) 质粒 pKMENGRbla 是由 pAK400 衍生而来的[26]。

(8) 限制性内切核酸酶 *Sfi*I 和 *Hin*dIII。

(9) 质粒 p KJE _ Bla-NΔ5：表达 TEM-1 β-内酰胺酶的 N 端前 5 个氨基酸残基缺失突变体（见 16.3.3）。

(10) Pr _ sfi _ pelB _ DG _ shuffl 的正向引物：5′-TTACTCACGGCCCAGCCG-GCCATGGCTGACGG-3′，有下画线的是限制性内切酶位点。

(11) Pr _ GG _ his5 _ hind _ shuffl 反向引物：5′-TAGTCAAGCTTACTAGT-GATGGTGATGGTGGCCACC-3′。

(12) Taq DNA 聚合酶（Sigma）。

(13) DNase I（Sigma）。

(14) DNase 缓冲液：50mmol/L Tris-HCl，pH7.5，1mmol/L $MgCl_2$。

(15) EDTA 溶液：0.5mol/L EDTA。

(16) 琼脂胶：1%～2%的琼脂溶在 0.5× Tris-base-boric 酸-EDTA（TBE）缓冲液。

(17) 10 × TBE 缓冲液：1mol/L Tris-base、1mol/L boric 酸、20mmol/L EDTA，pH8.0。

(18) Qiaex II 切胶回收试剂盒（Qiagen）。

(19) dNTP（Amersham）。

(20) 水浴锅（Epperndorf）。

(21) $MgCl_2$ 储液：25～100mmol/L $MgCl_2$。

(22) $MnCl_2$ 储液：5mmol/L $MnCl_2$。

(23) GFX PCR DNA 和切胶回收试剂盒。

16.2.2　转化、蛋白质表达和纯化

（1）丁醇。

（2）电感受态大肠杆菌 XL-1 Blue 细胞。

（3）Electroporator（Bio-Rad）。

（4）2YT：16g 胰化（蛋白）胨、10g 酵母提取物、5g NaCl 溶于 1L 水中，高压灭菌。

（5）转化盐储液：250mmol/L KCl 和 1mol/L $MgCl_2$。

（6）氨苄储液：将 100mg/ml 氨苄溶于水用 0.22μm 过滤器过滤，分装，−20℃储存。

（7）LB/Cm 固体板：1% 胰化（蛋白）胨、0.5% 酵母提取物、0.5% NaCl、1.5% 琼脂、25μg/ml 氯霉素。

（8）LB/Cm：1%胰化（蛋白）胨、0.5%酵母提取物、0.5% NaCl、1.5%琼脂、25μg/ml 氯霉素。

（9）2YT/Cm：2YT 培养基中加入 25μg/ml 氯霉素。

（10）高速离心机（Sorvall、GS-3 和 SS-34 转子）。

（11）悬浮缓冲液：50mmol/L 磷酸钠、500mmol/L NaCl，pH 7.0。

（12）Benzonase（Sigma）。

（13）0.45μm 磺酸多醚注射过滤器。

（14）2ml 苯硼酸柱子。

（15）硼酸缓冲液：0.5mol/L 硼酸、0.5mol/L NaCl，pH 7.0。

（16）4ml Ni-氨基三乙酸（NTA）柱子：4ml Ni-NTA 高流速介质（Qiagen）。

（17）咪唑缓冲液：50mmol/L 磷酸钠、0.25mol/L 咪唑、50mmol/L NaCl，pH 7.0。

（18）磷酸缓冲液：50mmol/L 磷酸钠、150mmol/L NaCl，pH 7.2。

（19）EDTA。

（20）1ml 苯氯仿 HR 5/5 柱子（Amersham-Pharmacia）。

（21）Tris-HCl 缓冲液：25mmol/L Tris-HCl、25mmol/L NaCl，pH 8.0。

（22）Mono Q HR 5/5 柱子（Amersham-Pharmacia）。

（23）Tris-HCl-NaCl 缓冲液：25mmol/L Tris-HCl、0.5mmol/L NaCl，pH 8.0。

16.2.3　酶活体系及尿素诱导的去折叠

（1）头孢尼特罗溶液：0.2mmol/L 头孢硝噻吩、50mmol/L 磷酸钾、0.5%二甲基亚砜，pH7.0。

（2）分光光度仪：至少能 1s 读 1 次数值，如，Ultrospec 3000（Amersham-Pharmacia）或 V-550（Jasco）。

（3）分光荧光计，如，FluoroMax-2（Jobin-Yvon）或 FP-6500（Jasco）。

（4）磷酸缓冲液［见 16.2.2（18）］。

（5）尿素：超纯，至少 99%纯度（ICN）。

16.3　方　　法

本章主要介绍提高蛋白质热稳定性的常规方法，而不需要在高温下筛选酶活性。介绍完 β-内酰胺酶模型系统（见 16.3.1），我们描述末端截切（见 16.3.2 和 16.3.3）

的设计方案，它们对酶活的影响（见16.3.4）。突变库由酶的定向进化和错误倾向PCR产生（见16.3.5），然后在体内应用不同的筛选压力进行筛选（见16.3.6）。在本节的最后一部分讨论截切突变体的再延长（见16.3.7）、表达策略和纯化方案、酶学方法（见16.3.9）以及稳定性检测（见16.3.10）。

16.3.1　β-内酰胺酶模型系统

为了说明截切-优化-再延长可以用来提高蛋白质的热稳定性，β-内酰胺酶被选作模型系统（图16.1A）。作为一个临床发病机制相关的研究因子和重要的抗体，它所属的家族已被广泛研究，多个晶体结构已解析。许多对该酶的突变已被用于更好地研究和解释结构-功能相关性以及理性设计截切片段。此外，稳定的β-内酰胺酶还被用于癌症治疗的潜在药物的活性研究，以便得到潜在的药物前体化合物[27,28]。

β-内酰胺酶的A家族（EC 3.5.2.6）是细菌胞质酶，其氨基酸序列多样化，稳定性不同，但其三维结构很相似。尽管β-内酰胺酶的A家族的肽主链在核心区能很好地重叠在一起，但在两个末端并没有很好的重叠性。末端结构的不同也与氨基酸长度和氨基酸组成相符合，让这类酶适合于研究序列同源和功能的相关重要性。

16.3.2　设计末端截切

缺失突变体的正确设计很重要：如果末端截切的过多，由于结构稳定性不够和错误折叠，其得到的截切突变体将没有功能；同样，如果截切的不够多，很有可能得不到期望的表型，选择压力将无法选择出稳定的结构。

如果目标蛋白的结构已被解析，可用Protein Data Bank（PDB；http：//www.pdb.org/），SWISSPROT（http：//www.expasy.org/sprot/和 http：//www.ebi.ac.uk/swissprot）以及SCOP（http：//scop.mrc-lmb.cam.ac.uk/scop/）等数据库以及结构分析工具，如分子模型程序WHATIF（http：//swift.cmbi.kun.nl/WIWWWI）和CE运算法（http：//cl.sdsc.edu/ce.html），来进行结构比较和分析。另外，接触图可用作识别可能的非重要和重要的相互作用，这些在设计过程也很重要。如果没有高分辨结构，可用其他的数据库（如HSSP）和SWISS-MODEL同源模型服务器（http：//swissmodel.expasy.org），或者用一个或两个氨基酸逐切来决定最大截切数。

在该研究中，根据以前的β-内酰胺酶的突变体研究[29]而设计4个截切突变体，对β-内酰胺酶的A家族进行结构同源序列比对，用CE运算法[30]和SWISS-PDB浏览器[31]（http：//www.expasy.org/spdbv）（图16.1B和图16.1C）来分析结构。截切的突变体试图在生理温度下引入结构干扰而不会使蛋白质完全没有功能。N端前3个氨基酸（His、Pro和Glu）有很高的温度因子（高达23.8Å^2，平均13.1 Å^2）和溶剂可接触性（PDB号1btl），当去掉前3个氨基酸得到的突变体NΔ3对蛋白质稳定性的影响比较小。NΔ5突变体去掉临近的苏氨酸和亮氨酸，其中苏氨酸被完全掩盖并与C端的Ser285氨基酸形成氢键。在C端，只有1个和3个氨基酸被去掉（突变体CΔ1和CΔ3），因为以前的研究表明末端酪氨酸是关键的氨基酸[32]。

16.3.3 载体质粒的设计

载体质粒被设计为当野生型β-内酰胺酶或截切的突变体与用于细胞间质定位的N端pelB信号序列[33]融合表达。一个短的天冬氨酸-甘氨酸的链接被引入用来确保N端不同的突变体同等的处理效率。最后，C端连接His-tag可用于固定金属离子亲和柱来纯化，甚至能够纯化非野生型的β-内酰胺酶变体（图16.2A）。

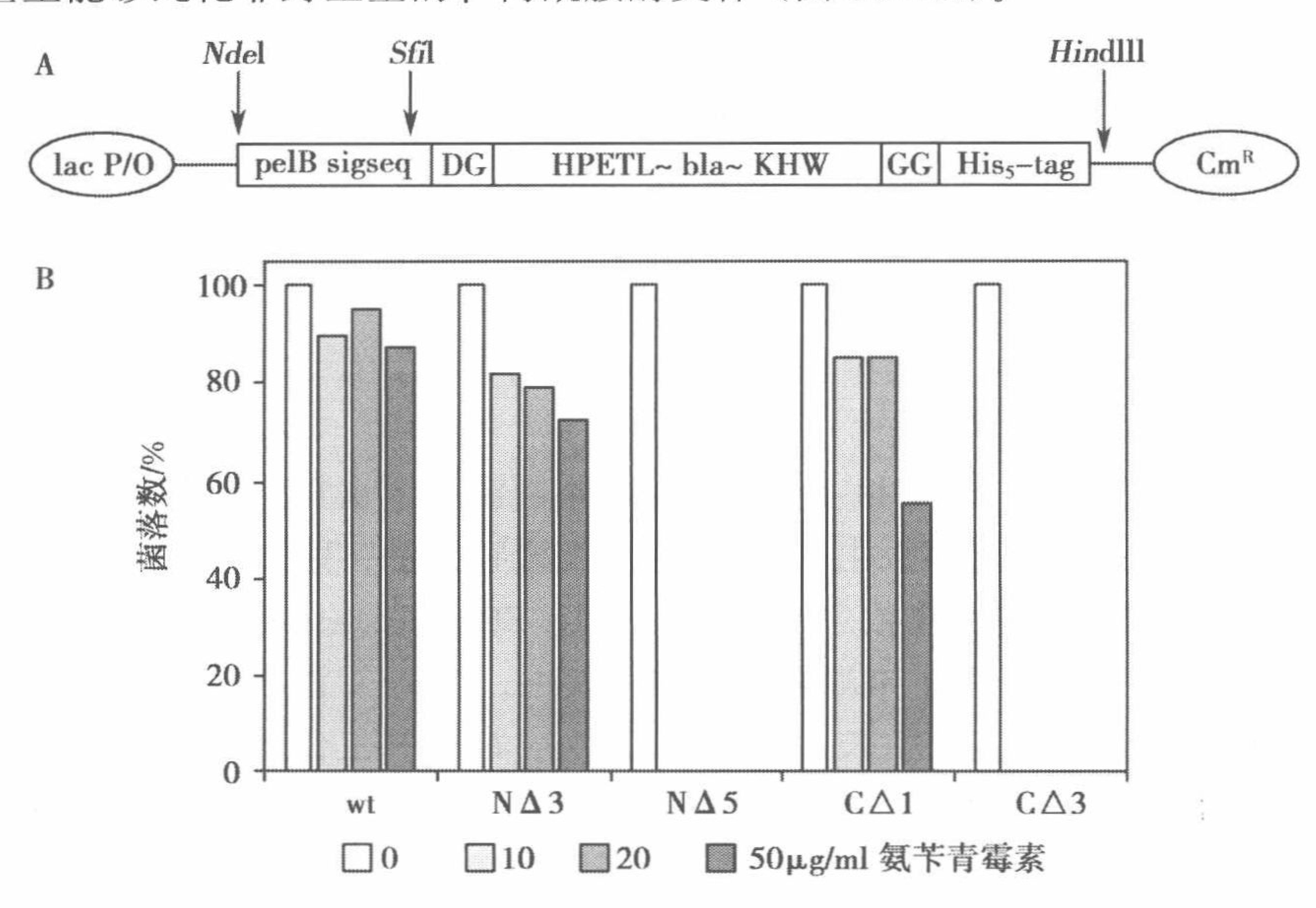


图16.2 A. 表达载体设计。在一段pelB的信号序列后面接着天冬氨酸和甘氨酸，之后是成熟的TEM β-内酰胺酶的编码区域，这个TEM β-内酰胺酶的后面是2个甘氨酸和1个用来辅助纯化用的His_5-tag。重要的限制性内切核酸酶位点被标记了出来。在细胞质表达的变体中，pelB信号序列和天冬氨酸/甘氨酸残基被1个单个的甲硫氨酸残基替代了。所有的质粒都包含1个氯霉素抗性基因。B. 在生物体中，适当的β-内酰胺酶最初缺失突变体。大约3500个新制的被转化的XL-1 Blue细胞带有相应的质粒，会在不同LB培养基平板上被删选，这些LB琼脂培养基分别含有氯霉素和10～50μg/ml的氨苄青霉素或者只有氯霉素（100%克隆）。平板在37℃下孵化20h后被分析。对于NΔ5克隆和CΔ3克隆，在氨苄青霉素存在时，没有克隆出现。

载体构建步骤如下：

（1）TEM-1β-内酰胺酶截切突变体NΔ3、NΔ5和CΔ1及CΔ3是由pUC衍生的*TEM*-1基因扩增而来[34]，分别用其正向和反向引物。正向引物含5′*Sfi*I酶切位点，还包含pelB信号肽链的C端编码序列和TEM-1截切突变体的N端修饰（截切）序列。在基因的3′端，包含反向序列编码改造的TEM-1的C端序列、1个短的序列（2个甘氨酸残基）、5个His-tag、2个终止密码子和*Hin*dIII酶切位点。

（2）其PCR产物由*Sfi*I/*Hin*dIII双酶切、纯化，然后克隆到pKMENGRbla（由K. M. M. 提供）载体中，pKMENGRbla是由表达载体pKA400[26]改造而来，含有1个氯霉素抗性基因。

得到的质粒（pKJE_Bla—NΔ/CΔ系列）在*lac*操纵子下表达了各种TEM-1突变体基因。尽管质粒上的*lac*I有组成型表达，强的T7g10 Shine-Dalgarno序列[35]有很高的表达水平，因此所有选择步骤在极端条件下进行，并未过表达。

16.3.4 末端截切对酶活性的影响

为了检测末端截切对酶活性的影响，在氨苄抗性递增的固体培养基和液体培养基中测其生长速度。通常，在固体板上的筛选比液体状态下更为苛刻。通过本底表达及IPTG 诱导表达以及在不同温度下的诱导来观察体系的不同。

(1) 如果有需要，用标准方法转化 BL21 细胞涂在 LB/Cm 板上。

(2) 从甘油菌或者转化的过夜（16h）培养的 LB/Cm 板上［步骤（1)］挑取单克隆。

(3) 将过夜培养的菌，在新培养基中稀释至 OD_{600} 为 0.1，并将其培养至 OD_{600} 为 1（见注 1)。

(4) 将预培养的菌液［步骤（3)］涂在含 IPTG 与不含 IPTG 的不同浓度的氨苄抗性板上。被涂的菌量可估量为 2.5×10^8 个细胞/ml，1 个 OD_{600}。

(5) 在液体 LB/Cm 培养基中加入步骤（3）的菌液，在含 IPTG 与不含 IPTG 的不同浓度的氨苄培养基中培养。

(6) 生长速率［步骤（4）和（5)］可以在不同温度下重复。

在不诱导情况下，于固体板上 37℃培养，当氨苄抗性浓度从 0μg/ml 增加到 50μg/ml 时，所有截切突变体的克隆数急剧减少（图 16.2B)。在连续培养 40h 后突变体 NΔ5 和 CΔ3 没有长出克隆，说明了末端氨基酸的重要性。当 25℃诱导时，NΔ5 能够在高达 50μg/ml 的氨苄固体培养基中生长，在 37℃液体培养基中在合适的供氧情况下仍能生长。这说明两个截切突变体都能干扰蛋白质结构，但同时可以折叠成有活性的结构。NΔ5 突变体比 CΔ3 对抗生素的容忍性更小，因此更适合于下一步的优化。

16.3.5 DNA 混编和随机突变

逆转由末端截切或删除突变造成的结构干扰的关键优化步骤是在遗传水平上建立高度可变的突变库。随机突变与 DNA 重组的结合能够满足这样的突变需求。

DNA 混编[15,16]是在试管里进行 DNA 重组的广泛应用的方法。由 DNA 混编方法将不同源的基因混杂在一起是定向进化中的关键步骤，其主要特点简述如下：在全基因上进行随机点突变后，主要是由易错 PCR（见 16.3.7）引入突变得到的突变库可用来 DNA 片段化，从而得到小的、略微不同的、可相互交换的、长度为 50～200 个碱基对的 DNA 片段。将一个或多个核苷酸不同的同源序列交迭并用 PCR 重新扩增成长片段，起初用低温退火来确保最小的片段也能够进行 PCR。随着循环次数的增加，将退火温度逐步提高来选择长的交迭片段。最后，用基因两侧的引物合成全长的基因，并引入合适的酶切位点进行下一步的克隆。

以下内容介绍用易错 PCR 和 DNA 混编来创建有足够复杂度、多样化的突变库。值得注意的是，这里介绍的步骤只针对我们的系统，随着目标基因的长度和组成需要有所改变。需要考虑在易错 PCR 引进突变之前反应的覆盖率，第一次随机突变，第二次组合是不能用来决定混编过程中的错误率的（见注 2)。

(1) 得到足够量（5～10μg）的目标基因，可用目标基因两端的同源序列 PCR 或者用限制性内切核酸酶在基因末端剪切。如果用 PCR 方法，应该用 *Taq* DNA 聚合酶来合成产物。在我们的例子中，我们以 pKJE _ Bla _ NΔ5 为模板，用引物 Pr _ sfi _ pelB _ DG _ shuffl 和 Pr _ GG _ his5 _ hind _ shuffl 扩增截切的 β-内酰胺酶。

(2) 将 4～5μg 的目标基因用 0.2U 的 DNase I 在 50μl DNase I 的酶切体系中，于室温（25℃）下酶切 12min。

(3) 加入 4μl 的 EDTA 溶液终止内切酶反应，并在冰上放置。

(4) 将得到的 DNA 片段在 2%的琼脂胶上分析并回收长度为 50～150bp 的 DNA 片段。

(5) 用 QiaexII 胶回收试剂盒回收 DNA 片段（见注 3）。

(6) 在 50μl 体系中，加入约 100～300μg 的纯化的片段用来作引物延伸 PCR（见注 4）。反应混合物包括 2.5U *Taq* DNA 聚合酶、0.4mmol/L dNTP2mmol/L $MgCl_2$。在 PCR 仪（Eppendorf）中，PCR 反应用以下程序进行：94℃ 3min；94℃ 1min（变性），(45 + 0.3)℃每个循环 1min（退火），72℃ 1min（延伸）（10 个循环）；最后一个循环后 72℃ 5min；94℃ 1min，(50+0.4)℃ 每个循环 1min，72℃ 1min（15 个循环）；最后一个循环后 72℃ 5min；94℃ 1min，(56+0.5)℃ 每个循环 1min，72℃ 1min（10 个循环）；最后一个循环后 72℃ 5min；94℃ 1min，(61+0.5)℃ 每个循环 2min，72℃ 2min（10 个循环）；最后一个循环后 72℃ 5min。另外，一个简单程序，如 94℃ 3min，35 个循环：92℃30s，(30+ 1)℃ 每个循环 1min，72℃ 1min + 4s/循环，最后一个循环后 72℃ 5min。

(7) 扩增 1/5 体积的重叠的 PCR 反应物（不需要纯化）来扩增全长基因，并加入合适的酶切位点以便下一步克隆。这步可在易错倾向条件下进行，以便扩增突变库的多样化。除了用 *Taq* DNA 聚合酶延长外，扩增反应的突变率可通过 7mmol/L $MgCl_2$、0.5mmol/L $MnCl_2$的加入而提高[36]（见注 5）。在我们的例子中，我们用引物 Pr _ sfi _ pelB _ DG _ shuffl 和 Pr _ GG _ his5 _ hind _ shuffl，每条引物的浓度是 0.5μmol/L。反应混合物包括 7mmol/L $MgCl_2$、0.5mmol/L $MnCl_2$、0.4mmol/L dNTP 和 2.5U *Taq*。所用的程序是 94℃ 3min，25 个循环：94℃ 1min，68℃ 1min，72℃ 1min，最后一个循环后 72℃ 7min。

(8) 用 GFX PCR DNA 和切胶回收试剂盒来纯化 PCR 产物，用合适的酶酶切产物，克隆到表达载体（pKJE _ Bla-NΔ5）中。用 16.3.6 所述方法进行转化。

我们进行了 3 个循环的定向进化（S1～S3），用 DNA 混编和随机突变并结合体内筛选过程。所得到的克隆被收集，作为下一步进化的模板。在定向进化的最后一个循环中，DNA 混编后的易错 PCR 被标准的 PCR 程序取代，防止在重组的克隆中产生有害突变。

16.3.6 转化及在体内筛选突变库

(1) 把质粒突变库［pKJE _ Bla-ΔN5 _ Lib-S1～S3，S 代表突变循环 1～3，（见 16.3.5)］，用丁醇沉淀。用 10～20μl 连接混合物，加入双倍量的去离子水使体积达到 50μl。加入 500μl 丁醇在室温离心 2000*g*，30min。去掉上清液，在室温晾干，加入 10～20μl 的水。

(2) 将沉淀后的 DNA 加入到 100μl 电转感受态 *E. coli* XL-1 蓝细胞中。1.7kV，200 Ω，25μF。转化完后立即加入 900μl 的 2YT 培养基和 1/100 体积的转化盐储液。

(3) 将悬浮的细胞加入 10ml 玻璃试管中，在 37℃培养 60～70min。

(4) 通过将转化产物逐步稀释涂在 LB/Cm 板上来检测转化效率。

(5) 用另一管转化的细胞涂在不同浓度的氨苄抗性 LB/Cm 板上来检测筛选压力（见注 6）。选择在最高氨苄浓度下长出来的突变体来进行下一步实验。

(6) 作为筛选，把转化的细胞涂在不同浓度的氨苄抗性 LB/Cm 板上［由步骤（5）得到的浓度］（见注 6）。我们用的氨苄在第一到第三次筛选的浓度分别为 20μg/ml、

100μg/ml、200μg/ml。第一次的筛选是在液体培养基中进行，后两次筛选是在固体培养基上。

突变体 NΔ5 经过 3 个循环的定向进化（包括随机突变和 DNA 混编）（见 16.3.5），被损功能得以恢复。在这 3 个循环中得到 19×10^3 个，147×10^3 个和 260×10^3 个的克隆，分别在 20μg/ml，100μg/ml 和 200μg/ml 氨苄抗性板上生长［步骤（6）］，得到 3600～7000 个克隆。

表 16.1 总结了所有在 3 个混编循环（S1～S3）的定向进化中得到的克隆的测序结果。另外，表 16.1 还包括了每个野生型残基的相对溶剂可接触性和平均侧链原子温度因子（由 PDB 1btl）[37]。其中在第二个循环后的 5 个突变体有两个突变，M182T 和 T265M（根据 Ambler et al. 命名）[38]，这两个突变在第一次循环后已经存在。

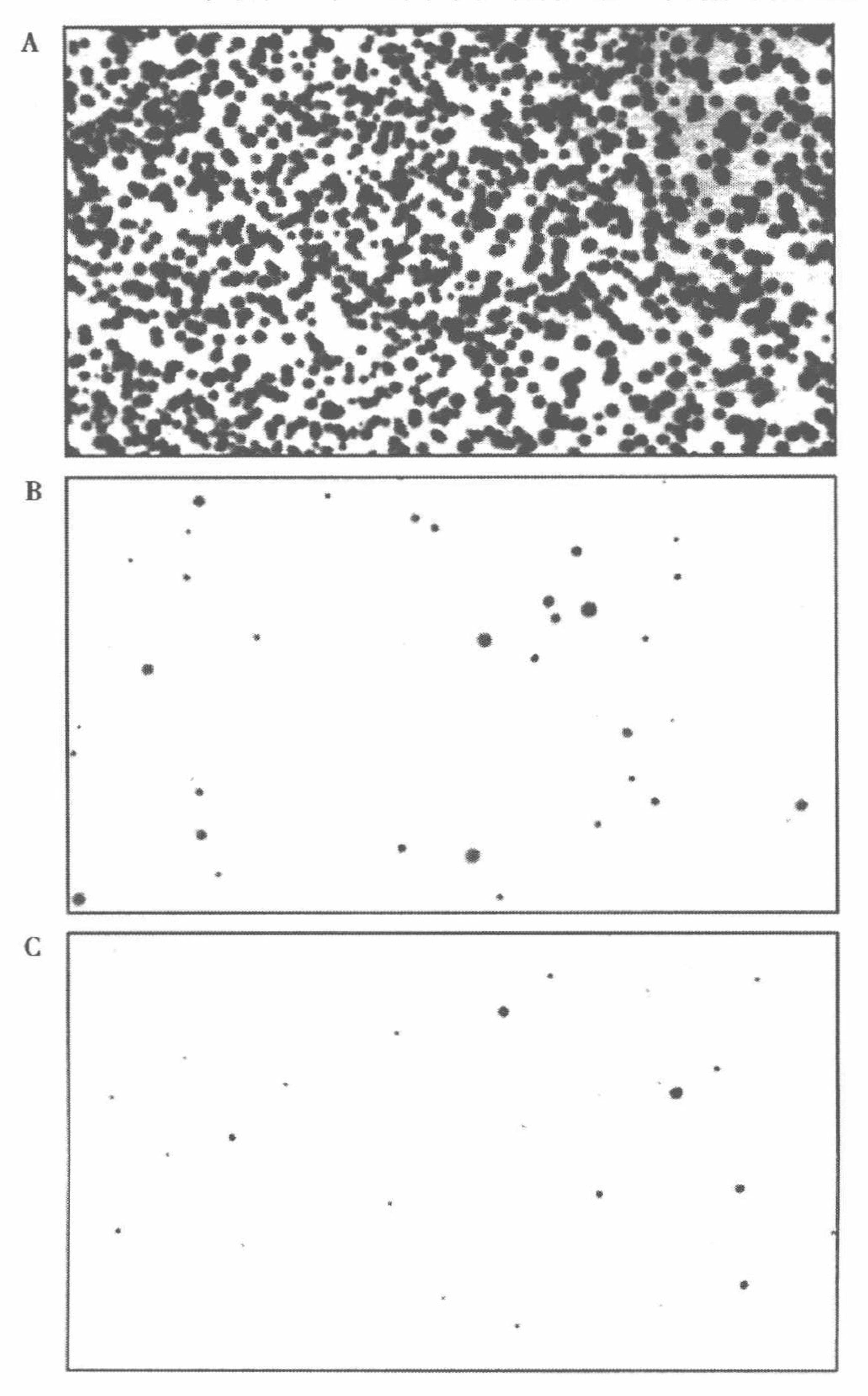


图 16.3　在 LB 琼脂平板培养基上优化的 NΔ5-β-内酰胺酶缺失突变体的克隆，培养基分别含有 A. 0μg/ml、B. 350μg/ml 和 C. 700μg/ml 氨苄青霉素。重新培养的在第三轮定向演化中分离的克隆，大约有 1500 个细胞，对其进行涂板。这些是平板在 37℃下孵化了 A. 20h，B 和 C. 40h 后拍摄的图片。

表 16.1 在 NΔ5 缺失突变体的定向进化中选择的氨基酸替换

position	31	37	38	42	52	56	59	63	82	84	88	96	104	120	147	150	153	159	168	177	182	192	195	198	206	208	224	240	247	256	265	277
wt	V	E	D	A	N	I	S	E	S	I	Q	H	E	R	E	A	H	V	E	E	M	K	T	L	Q	I	A	E	I	K	T	R
% acc	63.3	15.1	55.7	18	84.3	45.8	28.7	66.3	24	19.1	59.9	68.2	66.3	47.4	50.1	40.1	51.1	25.9	54.2	48.1	20.3	36	45.5	40.7	25.6	18.5	27.5	56.5	0	73.1	4.7	60.8
B factor(sc)	8.3	12.2	19.9	11.6	16.7	8.7	11.0	25.5	8.8	13.8	33.5	20.7	34.1	25.0	20.6	8.3	16.5	6.5	21.1	23.5	7.9	12.1	14.6	25.7	14.0	13.8	8.2	26.1	5.4	26.0	7.0	30.9
S1/1																			A		**T**										M	
S1/2													G								**T**		**S**				**V**			R		
S1/3																	**R**												V			
S2/1																	**R**				**T**				H		**V**				M	
S2/2															G					K	**T**										M	
S2/3																					**T**										M	
S2/4																			A		**T**	E	**S**				**V**				M	
S2/5				G																	**T**										M	
S3/1(×2)					S					**V**					G				V		**T**						**V**					
S3/2(×4)														G			**R**	A			**T**						**V**					
S3/4(×3)		D										Y									**T**						**V**				M	
S3/5(×4)	A													G							**T**						**V**		V			
S3/6																	**R**				**T**		**S**				**V**					
S3/7(×3)							G			**V**											**T**		**F**				**V**		V			
S3/17										**V**											**T**						**V**					
S3/18										**V**							**R**				**T**		**S**				**V**					
S3/19														G							**T**						**V**	G			M	K
S3/20	A		N	G					F	**V**											**T**		**S**	P			**V**		V	R	M	
S3/21	A																**R**				**T**		**S**				**V**					
S3/22									E		E						**R**		A		**T**						**V**			R		
S3/23					D			K		**V**						T	**R**				**T**		**S**	P			**V**					
S3/24						T				**V**							**R**				**T**					M	**V**					
S3/26																	**R**				**T**					M	**V**					
position	31	37	38	42	52	56	59	63	82	84	88	96	104	120	147	150	153	159	168	177	182	192	195	198	206	208	224	240	247	256	265	277

选择的缺失突变体的氨基酸替换相对于野生型序列的表述。残基根据 Ambler 等发表的文章编号[38]。这些突变体是根据在定向进化的哪一轮中选取的来分类(S1～S3)。一些出现频率很高的,包含相同的突变位点的克隆在括号中表示。那些在第三轮突变中最普遍的氨基酸取代用黑体字表示。% acc,相对来说溶剂已接近的程度,计算和真空之间的易接近程度的比例用百分比表示,通过程序 What If 计算得到[48];B 因子与来自蛋白质数据库(Protein Data Bank)的侧链原子的平均值相一致(PDB 号:1btl)。

经过最后优化步骤，收集突变库，在不用 IPTG 诱导的情况下，决定最大氨苄抗性浓度，见 16.3.4（图 16.3）。能长出显著数目的克隆数的抗性是 700μg/ml，经过 40 h 的培养，克隆可在含 1000μg/ml 的抗性上生长。相比，野生型只能在 500μg/ml 的抗性上生长。从含 800μg/ml 和 1000μg/ml 的抗性板上挑取 26 个克隆，并测序。26 个克隆属于 15 个不同的突变体（表 16.1）。所有的克隆有两个共同的突变位点，M182T 和 A224V。以前的研究表明 M182T 突变体能够互补折叠缺陷[39]和稳定性缺失[29,40]。突变体 A224V 在第三次循环中提高选择压力。这个突变体曾经被提到过，但并没有实验数据证实[41]。

两个突变体离截切的位点有 17Å 的距离，表明是由截切引起的结构干扰的独立互补。图 16.4 是对 TEM-1 β-内酰胺酶最常用的一些突变残基。通常，在单独优化的突变体中突变往往在整个结构中分布，而不是局部聚集在一起。

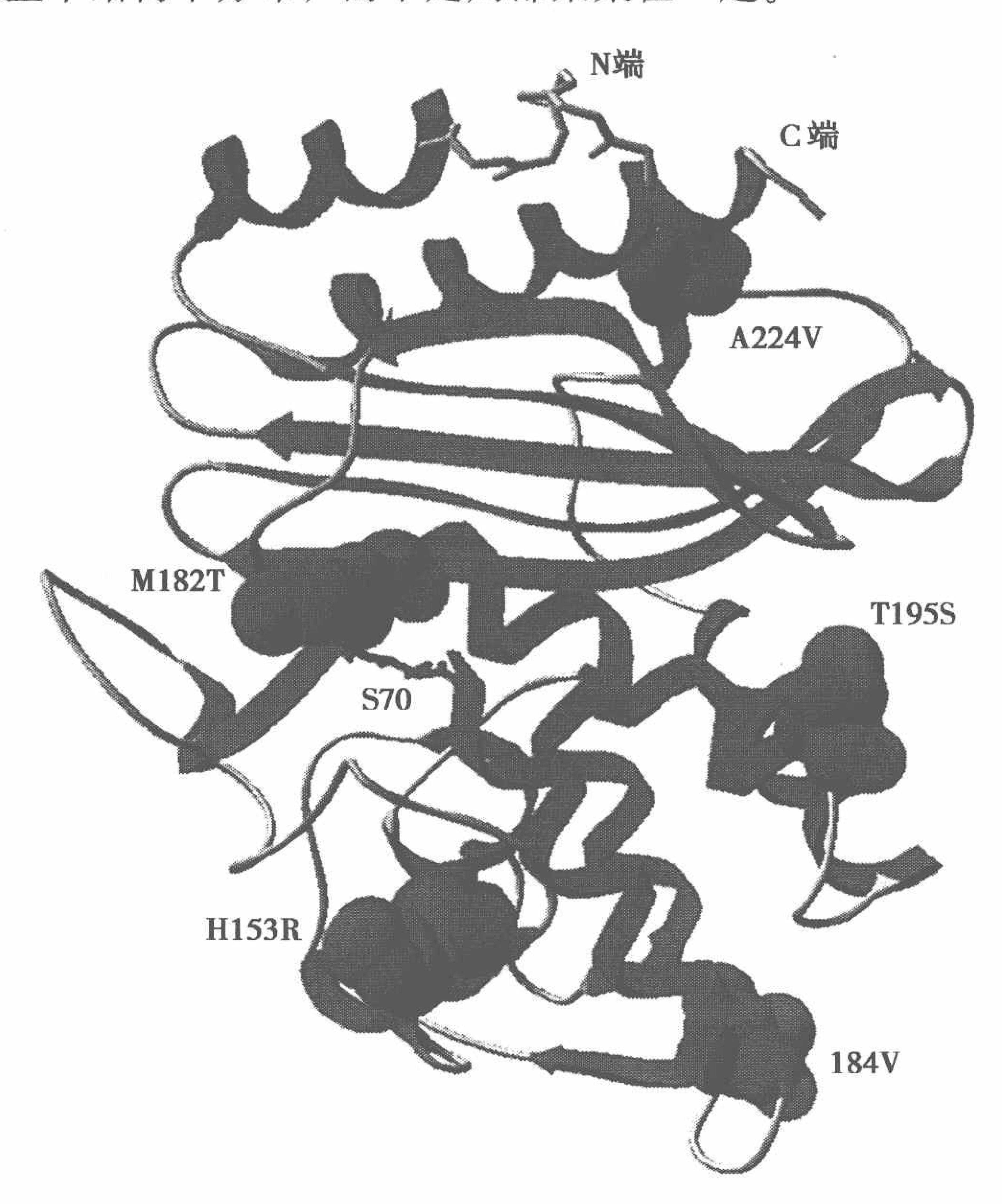


图 16.4　wt TEM-1 β-内酰胺酶（PDB 号：1bt1）的结构。经常突变的残基用全原子表示（如 M182T），催化位点丝氨酸残基（S70）以及 N 端和 C 端也被标明了。残基的编号与 Ambler 的编号一致[38]。大部分的突变氨基酸残基都位于蛋白质的表面附近。这张图片是用 SWISS-PDB Viewer 软件制作的[31]。

16.3.7　全长突变体的构建

优化的 NΔ5-S3/6 被再延长来研究突变对被截切后的不稳定蛋白质的互补作用。根据图 16.2A，被截切的氨基酸重新被加上，从而得到全长基因——被命名为 FL-S3/6 的克隆。这个突变体定位在细胞间质，在 *E. coli* XL-1 Blue 中表达来检测体内功能。让人意外的是，在含氯霉素和 100μg/ml 氨苄板上于 37℃ 20h 生长的克隆数目只有不含氨苄的对照

板上的50%。这和NΔ5-S3/6在两个板上得到相同数目克隆成鲜明对比。为了研究蛋白质功能和蛋白质定位，对FL-S3/6和NΔ5-S3/6的细胞粗提液做酶活反应。没有看到对发色底物头孢硝噻吩的水解反应的区别，说明酶的定位是很重要的。与这些数据相符合的是在FL-S3/6的纯化过程中也能看到信号肽剪切的减少（见16.3.8）。因此，可在细胞质内表达的FL-S3/6-cyt被克隆，该克隆用甲硫氨酸替代了信号肽以及Asp-Gly tag。为了保证在生理条件下正确的对比，与野生型内酰胺酶类似的基因（wt-clone-cyt）也被克隆。

16.3.8 蛋白质表达及纯化

β-内酰胺酶突变体（野生型间质定位的、野生型胞质定位的、NΔ5、优化的NΔ5-S3/6、NΔ5-S5/S7和FL-S3/6胞质定位的突变体）在lac启动子下于*E. coli*细胞中带His_5-tag融合表达（见16.3.8.1）。通过两步纯化反应进行纯化，首先用底物相似的亲和层析苯基硼酸盐柱子（见16.3.8.2），再用IMAC（见16.3.8.3）。这种纯化步骤非常适用于NΔ5（图16.5）和NΔ5-S3/7突变体的纯化。

对于某些β-内酰胺酶突变体（野生型、NΔ5-S3/6、FL-S3/6），其纯化的蛋白质有很高的污染并呈未加工完全的形态（30%～65%；见注7），极可能是因为过表达或者折叠太快造成的。由于即使是对旨在高产率的细胞间质提取物纯化也会引入污染，所以疏水亲和层析（HIC）被用来将蛋白质的自然形态与那些疏水性极高的未加工完全的蛋白质形态分离开来（见16.3.8.4）。

细胞质表达的β-内酰胺酶变体wt-clone-cyt和FL-S3/6-cyt用苯基硼酸盐亲和柱和IMAC进行纯化。可用离子交换柱进一步纯化（见16.3.8.5）。

16.3.8.1 蛋白质表达和细胞破碎

（1）将质粒转化到合适的表达菌株（如BL21）

（2）从甘油菌挑菌到2YT/Cm培养液中，在28℃培养16～18h（见注8）。

（3）用过夜菌扩培4个1L-2YT/Cm培养液，起初OD_{600}为0.15，在24℃培养NΔ5克隆，在25～30℃培养其他克隆（见注9）。

（4）OD_{600}为0.7时，用0.5mmol/L IPTG诱导细胞。诱导40min后，加入100μg/ml氨苄来筛选表达β-内酰胺酶的细胞。

（5）诱导4～5h后（OD_{600}为4～5），用GS-3转子收菌，在−80℃冻存。

（6）为了纯化，将菌（1L表达菌液）悬浮在悬浮缓冲液（见注10）中，加入200U的benzonase酶。

（7）在97MPa（约14 000psi）下反复循环6次破碎预冷细胞，要一直将样品放在冰上。

（8）将上清液用SS-34转子离心，41 000*g*，离心40min。用0.45μm磺酸多醚注射过滤器过滤上清液。

16.3.8.2 苯硼酸亲和层析柱

（1）将2ml苯硼酸盐柱用悬浮缓冲液悬浮，将16.3.8.1步骤（8）中得到的上清液上样到亲和柱中，用悬浮缓冲液冲洗直到280nm的吸收达到基线。

（2）用硼酸缓冲液洗脱β-内酰胺酶突变体。

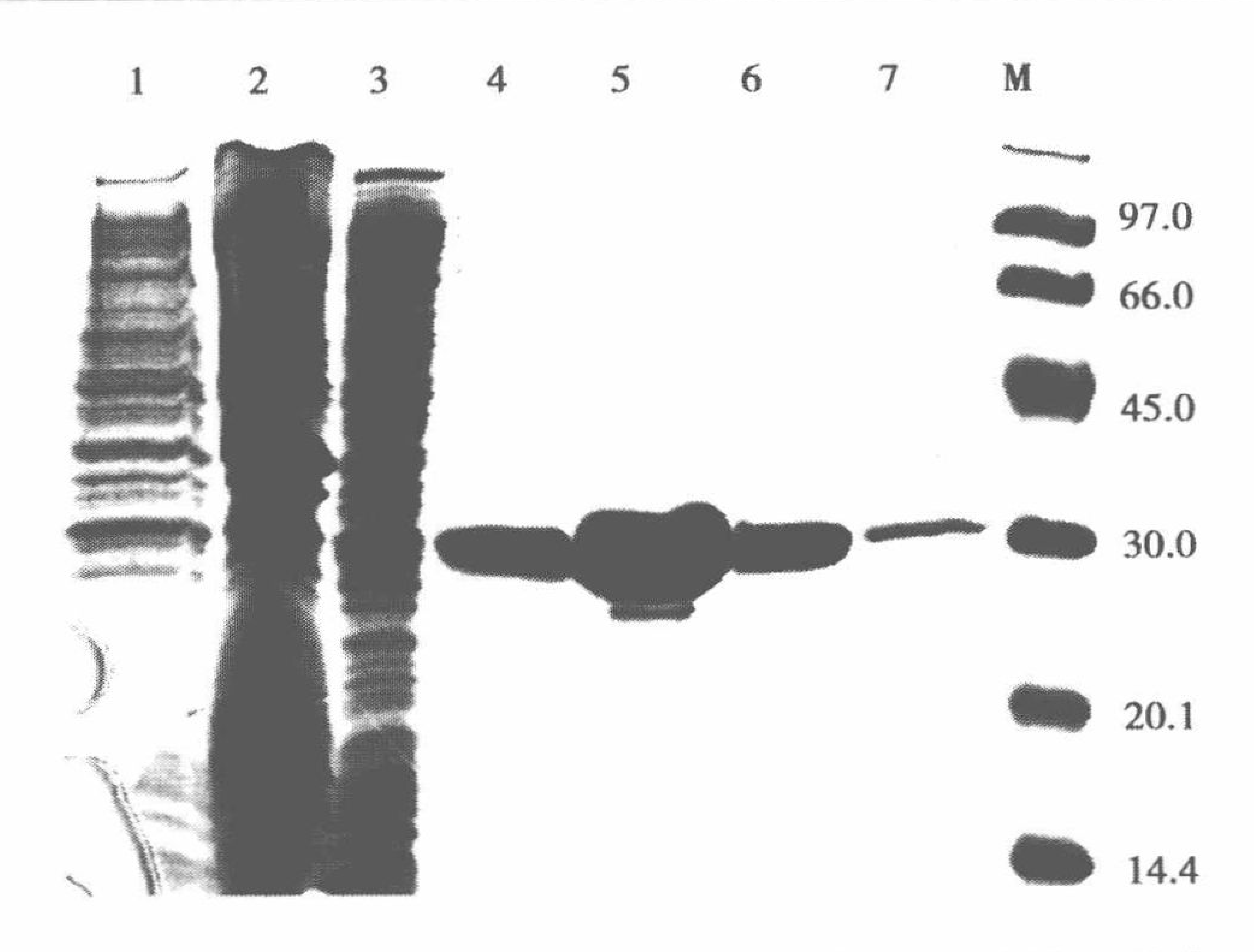


图 16.5　最初的β-内酰胺酶缺失突变体（NΔ5-克隆）的表达和纯化。12.5%的SDS-PAGE，考马斯（Coomassie）染色，展示了在分别用苯基硼酸盐（PheBo）和IMAC纯化前和纯化后收集的样品。第 1 列，诱导前的全细胞溶解产物；第 2 列，诱导后的全细胞溶解产物；第 3 列，细胞破碎后的上清液；第 4～7 列，IMAC洗脱峰的几个部分；M，分别标记了分子质量（kDa）的分子质量标签。

（3）用 12.5%的聚丙烯酰胺凝胶电泳（SDS-PAGE）来鉴定蛋白质纯度，用考马斯亮蓝染色。

16.3.8.3　金属亲和柱

（1）用含 5mmol/L 咪唑或不含咪唑的磷酸缓冲液平衡 4ml Ni-NTA 柱子，然后将 16.3.8.2 步骤（2）的样品上柱。

（2）用含咪唑分别为（5%、10%、16%和 100%）的梯度缓冲液洗脱样品。

（3）用 12.5%的聚丙烯酰胺凝胶电泳（SDS-PAGE）来鉴定蛋白质纯度，用考马斯亮蓝染色。

（4）为了进一步纯化，在含有 1mmol/L EDTA 的 1L 磷酸缓冲液中透析纯化的蛋白质（约 8 h），重复 3 次。

16.3.8.4　疏水作用色谱法

（1）用 1mol/L $(NH_4)_2SO_4$ 平衡 1ml 苯基硼酸盐柱。

（2）将 16.3.8.3 步骤（2）或 16.3.8.2 步骤（2）得到的样品加入硫酸氨到终浓度为 1mol/L $(NH_4)_2SO_4$（见注 11），离心（10min；27 000*g*；4℃；SS-34 转子）并过滤（0.45μm 过滤器）用以制备澄清样品。

（3）将样品上样到 HIC 柱子上，用 30ml 含 0～100% 0.5×磷酸的线性梯度缓冲液冲洗。疏水越强的部分与柱子结合越紧密。

（4）如果是苯基硼酸盐亲和层析后直接上 HIC 柱子，样品需要进一步用 IMAC 来纯化（见 16.3.8.3）。

16.3.8.5　离子交换层析

（1）将 16.3.8.3 步骤（2）中得到的样品在 1～1.5L Tris-HCl 缓冲液中透析 3～4 次。

（2）用 Tris-HCl 缓冲液平衡 Mono Q 离子交换柱并上样。

（3）用 30ml 0～100% Tris-HCl-NaCl 线性梯度 0～100%缓冲液洗脱。

（4）将洗脱得到的样品用节 16.3.8.3 步骤（4）所述方法透析。

所有突变体的氨基酸组成由电喷雾质谱法分析得以证实其符合计算质量的偏差范围。酶按照节 16.3.8.3 步骤（4）所述方法透析，并在纯化一周内进行鉴定。在用之前，酶液需要用离心方法（30min；15 000*g*；10℃）澄清，蛋白质浓度由在 280nm 处的吸收光谱测定（见注 12）。

16.3.9 酶活反应

β-内酰胺酶突变体酶活反应参数是根据发色底物头孢硝噻吩（nitrocefin，$\Delta\varepsilon_{486}$ = 16 000/(mol/L)/cm）在 486nm 的光谱来测定的。

（1）在 980μl 头孢硝噻吩溶液中加入 20μl 酶的稀释液，混合均匀后立即在 486nm 处测定在大约 1min 内的吸收值的变化（见注 13）。对于起始缺失突变，我们采用的酶的终浓度为 2.5nmol/L，对于其他变体则为 0.5nmol/L。

（2）上述测定过程被重复至少两次并计算标准偏差。

动力学常数是在 25℃下，在含 0.5%二甲基亚砜的 50mmol/L 的磷酸钾缓冲液（pH7.0）中，以 β-内酰胺化合物头孢硝噻吩为底物测定的。

16.3.9.1 米氏常数的测定及转化率的计算

（1）为了测定米氏常数（K_M），反应初速率是在底物浓度为 10～500μmol/L 时，用 2.7～5.9nmol/L 的 NΔ5 或 0.5nmol/L 其他变体来测量完成的。

（2）将数据输入到米氏方程，如用程序 SigmaPlot（SPSS）提供的 Marquardt-Levenberg 运算法来得到。

NΔ5 的 K_M 值与野生型相比没有变化，但是转化率 K_{cat} 降低了 6 倍，表明活性位点的构成对底物的结合是充分的，但并不能使有效的催化反应发生（表 16.2）。

表 16.2 动力学参量

β-内酰胺酶突变体	K_{cat}/（1/s）	K_M/（μmol/L）	K_{cat}/K_M/［mol/（L·s）］
野生型	780±9	84.4	9.24×10^6
NΔ5	123±2	61	2.02×10^6
NΔ5-S3/6	728±7	58.9	12.4×10^6
NΔ5-S3/7	732±19	64.3	11.4×10^6
野生型-cyt	602±8	52.2	11.5×10^6
FL-S3/6-cyt	633±9	70.6	8.97×10^6

16.3.9.2 热活性描述及半衰期

为了检测和比较对温度依赖的活性和热诱导失活的动力学参数，进行了热活性筛选。我们测定了在 25～70℃温度范围提高温度时反应的转化率（图 16.6）。

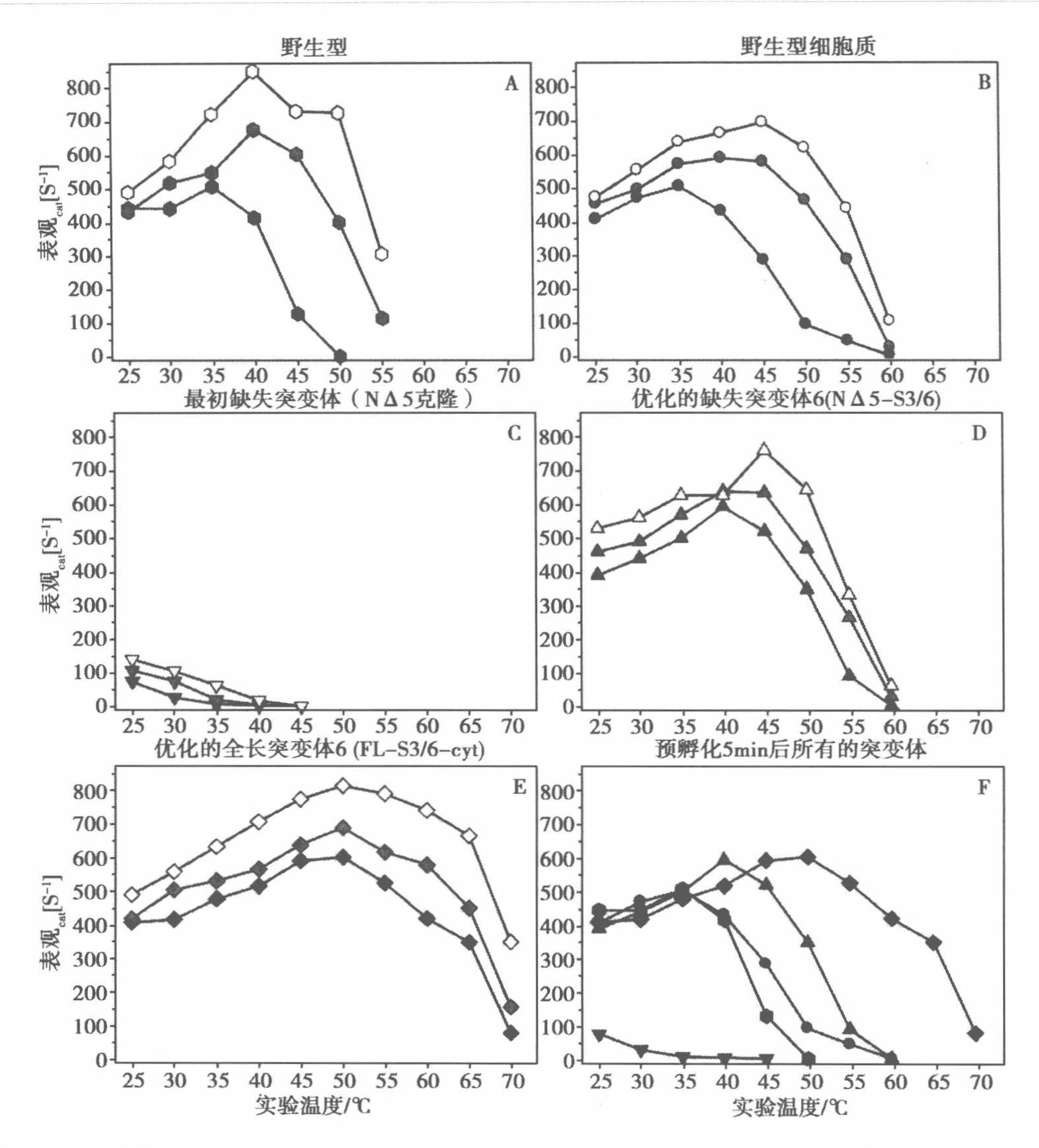


图 16.6　研究的β-内酰胺酶突变体的温度相关活性变化图。A～E. 表观 K_{cat}值是在 25～70℃范围内，以噻吩为底物，用分光光度计法测定的。酶溶液稀释（最初缺失突变体：292nmol/L，其他的突变体：25nmol/L），分成 3 个等份，并且进行不同的预处理：在实验温度下预热处理 5min 或者 30s（分别用黑色和灰色符号表示），或在冰上放置 10min（白色符号）。实验用缓冲液（980μl）先在热水浴中预加热，然后加入 20μl 酶溶液使反应开始。最后被利用的酶的浓度为 5.9nmol/L（最初缺失突变体）和 0.5nmol/L（其他的突变体）。F. 在预热处理 5min 后得到的热活性（thermoactivity）变化图的总结。⬢，野生型克隆（细胞间质）；●，野生型克隆（细胞液）；▼，NΔ5 克隆；▲，NΔ5-S3/6；◆，FL-S3/6-cyt。

(1) 为了检测热活性，配制 292nmol/L 的 NΔ5 突变体溶液及 25nmol/L 的其他突变体溶液，并分装成 3 等份。

(2) 将第一等份在已给定温度下（25～70℃）的水浴中放置 5min。其余两等份仍在冰上保存。

(3) 在检测之前，轻微的将样品混匀。

(4) 将 20μl 样品加入到 980μl 提前在反应温度下预热的头孢硝噻吩溶液中，反应

初速度在 5～25s 由加热的光谱仪测得。

（5）将第二等份在相应的温度下预热 30s，然后按照步骤（3）和（4）所述测酶活。

（6）把第三等份在冰上放置 10min 后测酶活。

（7）将上述步骤在不同温度下进行重复。

野生型的热活性主要依赖于预热处理（图 16.6A）。在冰上放置或预热 30s 的酶，其反应的最优温度是 40℃。然而，当酶预热 5min 后，其最优反应温度降至 35℃，这表明温度依赖的去折叠的出现。

NΔ5 突变体在 25℃（所检测的最低温度）下有最高酶反应活性（图 16.6C）。随着温度的升高，酶活迅速降低，当达到 40℃时，酶活为零。一个详细的在冰上温浴后在 40℃下测酶活实验表明，酶是在检测过程中被热失活的，而且发生在将酶加入反应液后的仅数秒内（图 16.7A）。随着温度的升高，酶活降低，通过指数衰减公式（见注 14），可算得半衰期在 40℃下为 7s（图 16.7B）。NΔ5-S3/6 突变体在 0s 和 30s 的热活性与野生型很相似。然而，5min 预热后的情况则很不一样。与野生型相比，NΔ5-S3/6 突变体的热活性提高大约 8℃（图 16.6A 和 F）。

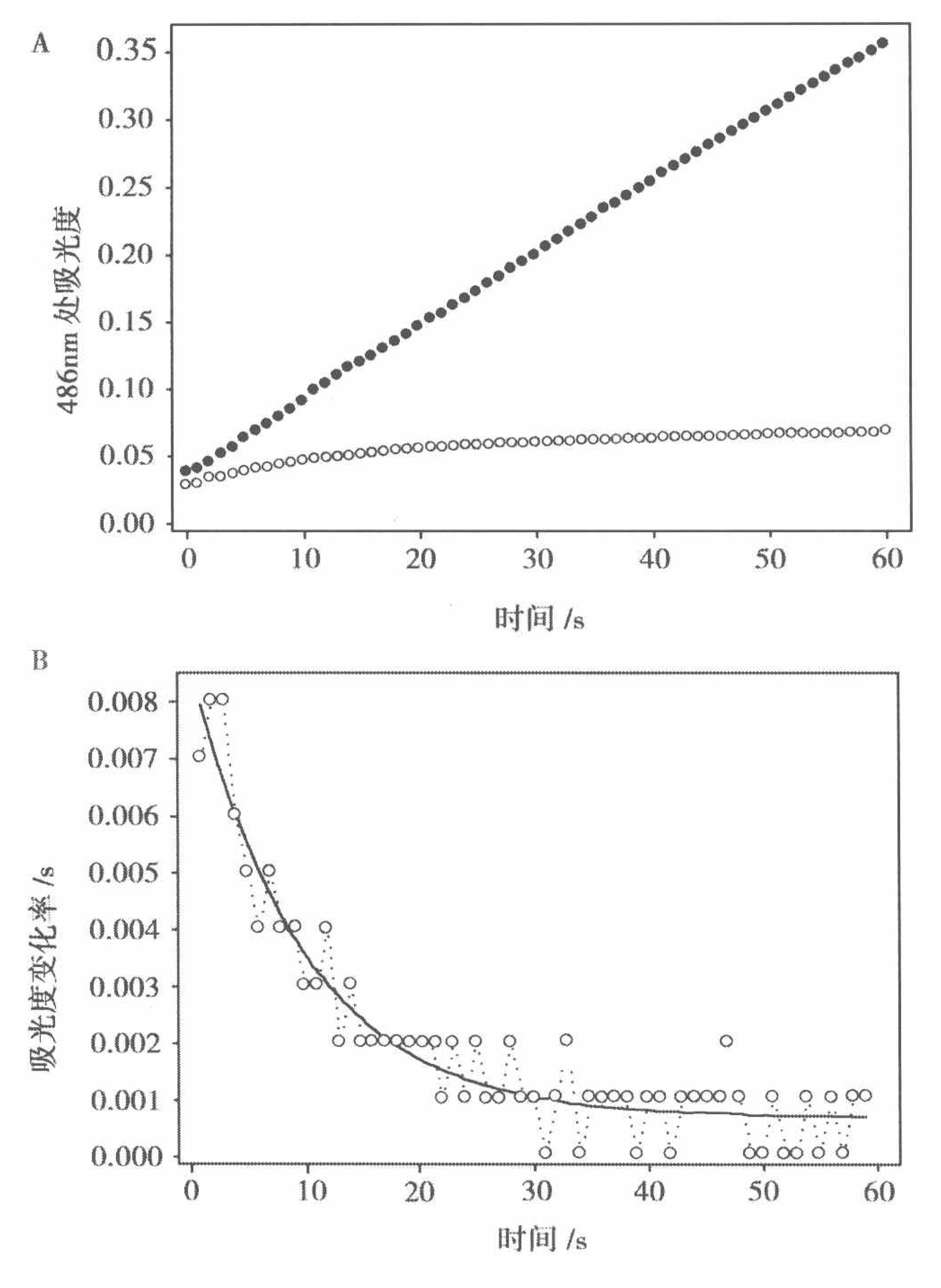


图 16.7 在 40℃时，酶催化动力学以及 NΔ5 克隆的催化活性指数衰减和 NΔ5-S3/6 的稳定动力学。A. 利用发色的底物硝噻吩来指示产物生成的光谱分析。酶的终浓度为 2.9nmol/L（NΔ5 克隆○）和 0.5nmol/L（NΔ5-S3/6 克隆●）。反应是将 20μl 每种浓度的酶溶液与 980μl 预加热的实验缓冲液混合后进行的。B. NΔ5 克隆的酶活性的指数衰减拟合，通过把 A 中的数据转换成每秒的吸收值（ΔA/Δt）。一个三参数的指数衰减拟合表明 NΔ5 克隆在 40℃时具有 7s 的半衰期（利用 SigmaPlot）。

全长的优化突变体 FL-S3/6-cyt（图 16.6E）展示了比优化缺失突变体 NΔ5-S3/6 以及相应的野生型 wt-clone-cyt 更好的热稳定性（图 16.6B 和 F）。这些酶的活性在 40℃以内很相似，而与预热与否关系不大。在温度高于 50℃时，FL-S3/6-cy 保持了显著较高的酶反应活性，尤其在延长热压力到预热 5min 的情况下。在冰上放置的野生型胞质内酰胺酶的最佳催化活性是在 45℃下，而预热 5min 后则降到 35℃。与之相比，FL-S3/6-cyt 的最佳温度（50℃）则保持不变。只是温浴后的反应速率有所下降。有意思的是，野生型和胞质野生型内酰胺酶的 N 端氨基酸（Met 取代 AspGly）的转换影响着酶的稳定性。

对所有截切突变体和全长突变体的 5min 预热后的热活性的比较（图 16.6F）阐明了末端截切是怎样在 35～40℃使活性减小的。补偿氨基酸后能够使活性得以恢复，甚至提高稳定性。经过再延长优化了的 S3/6 突变体进一步提高了酶的热稳定性和热活性。

16.3.10　尿素诱导的去折叠和数据分析

β-内酰胺酶突变体的去折叠是通过荧光分光光度计来检测的。色氨酸固有的最大荧光发射值的红移作为与尿素浓度相关的函数来测定（图 16.8，表 16.3）。根据三态模型进行数据分析。

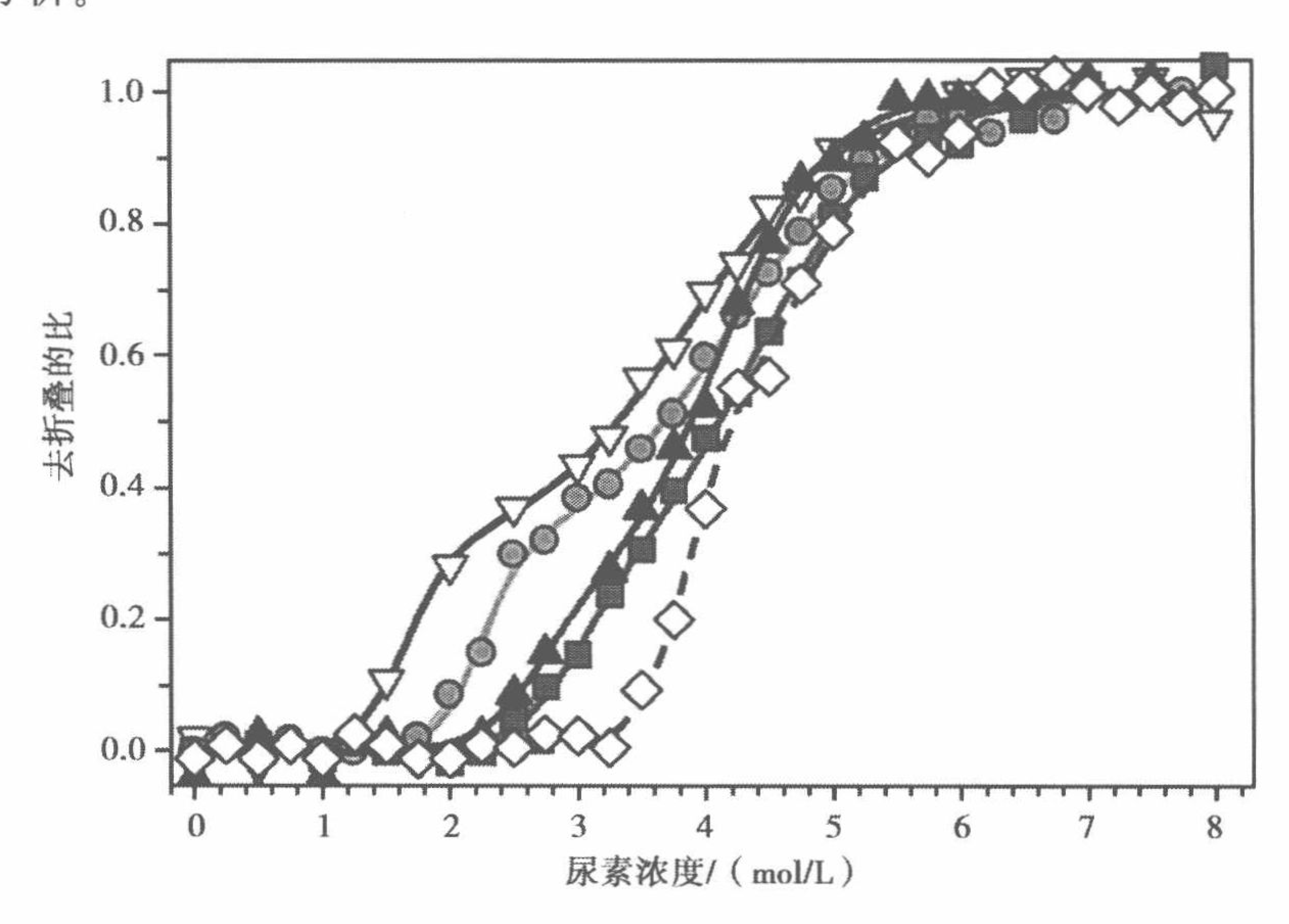


图 16.8　β-内酰胺酶在尿素中的去折叠。变性是通过固有的荧光发射最大值的红移来跟踪的。酶样品（0.3～0.4mmol/L，在 50mmol/L 磷酸钠和 150mmol/L 氯化钠中，pH7.2）含有 0.25～8mol/L 的尿素，在 19℃下平衡 18～20h，然后在 20～23℃下测定。数据是在假设为三态去折叠机制下拟合的。NΔ5 克隆用▽（白色）表示的；野生型克隆-cyt 用●（亮灰色）表示；优化缺失突变体 NΔ5-S3/7 用■（深灰色）表示；突变体 NΔ5-S3/6 用▲（黑色）表示；优化的全长突变体 FL-S3/6-cyt 用◇（白色，虚线）表示。

（1）将 300～400nmol/L 的酶溶液［见 16.3.8.3（4）］用 0.25～8mol/L 的尿素在 19℃下平衡 18～20h。

（2）在 20～23℃，记录 320～380nm 的荧光发射波谱。对每个数据点来说，平均扫描 4 次。

(3) 如果需要，则根据溶液（含有变性剂的磷酸缓冲液）背景荧光来校正荧光光谱。

表 16.3 热力学参量

突变体	$\Delta G^0_{NI,H_2O}$ /(kJ/mol)	m_{NI} /[kJ/mol/(mol/L)]	$D^{1/2}_{NI}$ /[/(mol/L)]	$\Delta G^0_{IU,H_2O}$ /(kJ/mol)	m_{IU} [kJ/mol/(mol/L)]	$D^{1/2}_{IU}$ [/(mol/L)]
Wt-clone-cyt	30.4±4.5	13.7±2.1	2.9	15.8±1.1	3.8±0.2	4.2
NΔ5	20.2±5.5	12.3±3.7	1.6	15.3±1.7	3.9±0.4	3.9
NΔ5-S3/7	27.1±6.4	9.3±2.6	2.9	19.0±2.6	4.3±0.5	4.4
NΔ5-S3/6	26.2±10.1	9.8±4.3	2.7	23.4±3.2	5.7±0.6	4.1
FL-S3/6-cyt	45.8±7.7	11.9±2.2	3.8	27.2±7.0	5.6±1.2	4.9

注：数据是用线性外推法分析的[49]，假设存在两对去折叠转变（自然态、中间态和去折叠态）。一半变性的值是根据方程（2）得出的［16.3.10（7）］。

(4) 从荧光强度谱中算出 λ_{max} 值。我们运用了程序 SigmaPlot（见注 15）中的权重功能（tribute weighting）来完成此计算。

(5) 分析假定得到的合适去折叠模型的数据。

(6) 对于一个三态模型，NIU（N 表示自然态，I 表示中间态，U 表示去折叠态）有两个平衡常数，K_{NI} 和 K_{IU}，利用下面的公式，可以以热力学参量 $\Delta G^0_{NI,H_2O}$、$-\Delta G^0_{IU,H_2O}$、m_{NI}、m_{IU} 和 y_I 为参数拟合实验得到的荧光数据（见注 16）：

$$y_{obs}=\frac{y_N+y_I\times e^{\left(\frac{-\Delta G^0_{NI,H_2O}+m_{NI}\times[D]}{R\times T}\right)}+y_U\times e^{\left(\frac{-\Delta G^0_{NI,H_2O}+m_{NI}\times[D]}{R\times T}\right)}\times e^{\left(\frac{-\Delta G^0_{IU,H_2O}+m_{IU}\times[D]}{R\times T}\right)}}{1+e^{\left(\frac{-\Delta G^0_{NI,H_2O}+m_{NI}\times[D]}{R\times T}\right)}+e^{\left(\frac{-\Delta G^0_{NI,H_2O}+m_{NI}\times[D]}{R\times T}\right)}\times e^{\left(\frac{-\Delta G^0_{IU,H_2O}+m_{IU}\times[D]}{R\times T}\right)}} \tag{1}$$

分别取 y_N 和 y_U 为第一个和最后一个数据点的平均值。在我们这种情况下，在基线处的一个倾斜没法看到。

(7) 转换的中点可以这样给出：

$$[D]_{1/2}=\frac{\Delta G^0_{H_2O}}{m} \tag{2}$$

蛋白质的变性（图 16.8，见注 17）与热活性（图 16.6，16.3.9.2）是相吻合的。稳定性最不好的是 NΔ5 突变体，然后是野生型胞质内酰胺酶。优化的突变体 NΔ5-S3/6 和 NΔ5-S3/7 在相同的尿素浓度下开始去折叠，被延长的 FL-S3/6 突变体是最后一个去折叠的。

野生型胞质内酰胺酶和 NΔ5 的去折叠表现了很明显的三态去折叠。FL/S3-6-cyt 突变体数据也表现了很好的三态去折叠。NΔ5-S3/6 和 NΔ5-S3/7 突变体数据既可以用二态，也可以用三态去折叠来解释。相比而言，所有的数据都能用三态模型 NIU，式（1）（步骤（6），表 16.3）来解释。这种方法由以前 TEM-1 β-内酰胺酶去折叠中间态折叠模型支持[42~44]。尿素处理后第一次变性的酶显示了降低的活性。从中得到的热力学值可以帮助比较和解释所得的突变体，但需谨慎使用，因为这是在假设的基础上得到的。去掉前 5 个 N 端的氨基酸（主要影响第一过渡态），稳定性被降低 10kJ/mol，对第二过渡态的影响不大。催化活性的体内优化得到对两个过渡态都有稳定作用的突变体，所得突变体在第一过渡态的稳定性在野生型之下，而第二过渡态的稳定性则在其之上。与 NΔ5-S3/6 相比，延长的 FL-3/6-cyt 明显稳定了第一阶段（19.6kJ/mol），

但对第二阶段的影响不大（3.8kJ/mol）。这与延长补偿截切突变体不受引入突变的干扰的现象是相符合的。与野生型胞质内酰胺酶 wt-clone-cyt 相比，FL-S3/6-cyt 的稳定性在第一过渡态增加了 15kJ/mol，在第二次过渡增加了 11.4kJ/mol。需要注意的是优化的突变体与野生型相比在更高的尿素浓度下才会变性，证实了在热活性实验中观察到的稳定性排序。

16.4 注

1. 为了确保生长结果可以比较，需要有一个统一起始条件的预菌液。

2. 经过分析 2529 个碱基，混编过程的错误率是 0.83。

3. 在延长温浴时间（15min）的情况下，洗脱的结合的 DNA 片段随着温浴温度的升高而增加，如 50℃。

4. 不需要额外引物来辅助分离的 DNA 片段之间互为引物并混编。

5. Mg^{2+} 的增加能稳定没有配对的碱基对，然而 0.5mmol/L Mg^{2+} 会降低聚合酶对模板的亲和力。错误率可以在广泛的范围内得到控制。例如，通过改变模板分子数、循环次数、延伸的时间、聚合酶的来源及用量、dNTP 的不同比例以及 dNTP 类似物的采用等。错误率仍然与目标基因的核苷酸组成是相关的。因此同样的操作在不同的实验中会得到完全不同的结果。

6. 如果第一次循环后酶活反应很低，可在溶液状态（LB/Cm 加入不同浓度的氨苄）下进行筛选，因为溶液的筛选压力比固体培养基的要少。

7. 未加工完全的蛋白质百分比可通过扫描考马斯亮蓝染色的胶，用 Scion/NIH 成像系统来定量。

8. 如果在低于 37℃下表达，最好把过夜培养菌也在相对较低的温度下培养以避免出现表达培养时的滞后现象。

9. 优化表达温度可通过对各个突变体在不同的温度下培养少量培养物来调整。

10. 在两个 pH（7.0 和 7.2）下检测，得到相同的结果。

11. 1mol/L $(NH_4)_2SO_4$ 可确保正确的结合，但是如果蛋白质容易聚集，可以适当降低 $(NH_4)_2SO_4$ 浓度。因为在野生型 β-内酰胺酶中加入 0.5mol/L $(NH_4)_2SO_4$ 时就开始沉淀，所以 $(NH_4)_2SO_4$ 的最终使用浓度为 0.65mol/L。在这个实验中，自然的蛋白质存在于穿出峰中，而没有加工完全的蛋白质则保留在柱子上。

12. 吸收谱测定的范围是 350～220nm。在 280nm 的背景吸收是通过用 320nm 和 350nm 之间的吸收与 280nm 的吸收线性外推来校正的。摩尔消光系数和分子量是根据 Gill 和 Hippel 来计算得到的[45]。

13. 对于半衰期短的酶来说，反应的头几秒非常重要，所以加入酶后应立即检测。另外，应该使用每秒至少能读一次数据的分光光度仪。

14. 酶学反应的衰减是利用 SigmaPlot 中的三参数指数衰减来拟合的：$\frac{\Delta A}{\Delta t}=a+\left[\frac{\Delta A}{\Delta t}\right]_0 e^{-\lambda\times t}$，这里 a 是一个说明硝噻吩溶液最初和积累吸收的一个参数，半衰期 $T_{1/2}=\ln 2/\lambda$。

15. 根据每纳米所需要的数据数目，各种参数需要被调节。另一种方法是用滑动平

均数来平滑数据。除了要确定 λ_{max}，我们也测定了 330～370nm 的重心迁移。尽管用了更多的数据，但这种方法并不一定能给出更好的结果。

16. 最大发射值（y_{obs}）是关于变性浓度（[D]）的函数，根据 3 个构象状态（y_N、y_I、y_U）的出现比率（f_N、f_I、f_U），把它们的组合信号进行去卷积化，可以用公式描述，即 $y_{obs}=y_N\times f_N+y_I\times f_I+y_U\times f_U$。质量作用定律 $K_{NI}=[I]/[N]$，$K_{IU}=[U]/[I]$，满足质量总和的守恒，$[N]_0=[N]+[I]+[U]$，这样可以计算它们的相对比例，$f_N=1/(1+K_{NI}+K_{NI}\times K_{IU})$，$f_I=K_{NI}/(1+K_{NI}+K_{NI}\times K_{IU})$，$f_U=K_{NI}\times K_{IU}/(1+K_{NI}+K_{NI}\times K_{IU})$。热力学参量建议运用线性指数方法，在 $\Delta G^0=-RT\ln K=\Delta G^0_{H_2O}-m\times[D]$ 的基础上，假设对所有的状态有一个线性关系[46]。

17. 为了说明两个荧光剂之间绝对最大值的微小变化，标准化的部分去折叠。图标中给出了公式：$f_{unfold}=(y_{obs}-y_F)/(y_U-y_F)$。

参考文献

1. Jaenicke, R. and Böhm, G. (1998) The stability of proteins in extreme environ- ments. *Curr. Opin. Struct. Biol.* 8, 738-748.
2. Ladenstein, R. and Antranikian, G. (1998) Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv. Biochem. Eng. Biotechnol* 61, 37-85.
3. Querol, E., Perez-Pons, J. A., and Mozo-Villarias, A. (1996) Analysis of protein conformational characteristics related to thermostability. *Protein Eng.* 9, 265-271.
4. Vieille, C. and Zeikus, G. J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* 65, 1-43.
5. Pace, C. N., Grimsley, G. R., Thomson, J. A., and Barnett, B. J. (1988) Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J. Biol. Chem.* 263, 11820-11825.
6. Mason, J. M., Gibbs, N., Sessions, R. B., and Clarke, A. R. (2002) The influence of intramolecular bridges on the dynamics of a protein folding reaction. *Biochemistry* 41, 12093-12099.
7. Thompson, M. J. and Eisenberg, D. (1999) Transproteomic evidence of a loop deletion mechanism for enhancing protein thermostability. *J. Mol. Biol.* 290, 595-604.
8. Vogt, G., Woell, S., and Argos, P. (1997) Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* 269, 631-643.
9. Szilagyi, A. and Zavodszky, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure Fold Des* 8, 493-504.
10. Karshikoff, A. and Ladenstein, R. (2001) Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends Biochem. Sci.* 26, 550-556.
11. Petsko, G. A. (2001) Structural basis of thermostability in hyperthermophilic pro-teins, or "there's more than one way to skin a cat." *Methods Enzymol.* 334, 469--478.
12. Malakauskas, S. M. and Mayo, S. L. (1998) Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* 5, 470-475.
13. Filikov, A. V., Hayes, R. J., Luo, P., et al. (2002) Computational stabilization of human growth hormone. *Protein Sci.* 11, 1452-1461.
14. Lehmann, M., Loch, C., Middendorf, A., et al. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* 15, 403-411.
15. Stemmer, W. P. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389-391.
16. Stemmer, W. P. (1994) DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* 91, 10747-10751.
17. Yang, F., Cheng, Y., Peng, J., Zhou, J., and Jing, G. (2001) Probing the conformational state of a truncated

staphylococcal nuclease R using time of flight mass spectrometry with limited proteolysis. *Eur. J. Biochem.* 268, 4227-4232.

18. Van der Schueren, J., Robben, J., and Volckaert, G. (1998) Misfolding of chloramphenicol acetyltransferase due to carboxy-terminal truncation can be corrected by second-site mutations. *Protein Eng.* 11, 1211-1217.

19. Sherwood, L. M. and Potts, J. T., Jr. (1965) Conformational studies of pancreatic ribonuclease and its subtilisin-produced derivatives. *J. Biol. Chem.* 240, 3799-3805.

20. Haruki, M., Noguchi, E., Akasako, A., Oobatake, M., Itaya, M., and Kanaya, S. (1994) A novel strategy for stabilization of *Escherichia coli* ribonuclease HI involving a screen for an intragenic suppressor of carboxyl-terminal deletions. *J. Biol. Chem.* 269, 26904-26911.

21. Trevino, R. J., Tsalkova, T., Kramer, G., Hardesty, B., Chirgwin, J. M., and Horowitz, P. M. (1998) Truncations at the NH2 terminus of rhodanese destabilize the enzyme and decrease its heterologous expression. *J. Biol. Chem.* 273, 27841-27847.

22. Trevino, R. J., Gliubich, F., Berni, R., et al. (1999) NH_2-terminal sequence truncation decreases the stability of bovine rhodanese, minimally perturbs its crystal structure, and enhances interaction with GroEL under native conditions. *J. Biol. Chem.* 274, 13938-13947.

23. Vainshtein, I., Atrazhev, A., Eom, S. H., Elliott, J. F., Wishart, D. S., and Malcolm, B. A. (1996) Peptide rescue of an N-terminal truncation of the Stoffel fragment of taq DNA polymerase. *Protein Sci.* 5, 1785-1792.

24. Shortle, D. and Lin, B. (1985) Genetic analysis of staphylococcal nuclease: identi-fication of three intragenic "global" suppressors of nuclease-minus mutations. *Genetics* 110, 539-555.

25. Petrounia, I. P. and Arnold, F H. (2000) Designed evolution of enzymatic proper-ties. *Curr. Opin. Biotechnol.* 11, 325-330.

26. Krebber, A., Bornhauser, S., Burmester, J., et al. (1997) Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Methods* 201, 35-55.

27. Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P., and Stevens, R. C. (2001) Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat. Struct. Biol.* 8, 238-242.

28. Rodrigues, M. L., Presta, L. G., Kotts, C. E., et al. (1995) Development of a humanized disulfide-stabilized anti-p185HER2 Fv-beta-lactamase fusion protein for activation of a cephalosporin doxorubicin prodrug. *Cancer Res.* 55, 63-70.

29. Huang, W. and Palzkill, T. (1997) A natural polymorphism in beta-lactamase is a global suppressor. *Proc. Natl. Acad. Sci. USA* 94, 8801-8806.

30. Shindyalov, I. N. and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 11, 739-747.

31. Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714-2723.

32. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P. S., and Palzkill, T. (1996) Amino acid sequence determinants of beta-lactamase structure and activity. *J. Mol. Biol.* 258, 688-703.

33. Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J., and Wilcox, G. (1987) Characterization of the *Erwinia carotovora pelB* gene and its product pectate lyase. *J. Bacteriol.* 169, 4379-4383.

34. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vec-tors. *Gene* 33, 103-119.

35. Olins, P. O., Devine, C. S., Rangwala, S. H., and Kavka, K. S. (1988) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene* 73, 227-235.

36. Cadwell, R. C. and Joyce, G. F. (1994) Mutagenic PCR. *PCR Methods Appl.* 3, S136-S140.

37. Jelsch, C., Mourey, L., Masson, J. M., and Samama, J. P. (1993) Crystal structure of Escherichia coli

TEM1 beta-lactamase at 1.8 A resolution. *Proteins* 16, 364-383.

38. Ambler, R. P., Coulson, A. E, Frere, J. M., et al. (1991) A standard numbering scheme for the class A beta-lactamases. *Biochem. J.* 276 (Pt 1), 269, 270.

39. Sideraki, V., Huang, W., Palzkill, T., and Gilbert, H. F. (2001) A secondary drug resistance mutation of TEM-1 beta-lactamase that suppresses misfolding and aggregation. *Proc. Natl. Acad. Sci. USA* 98, 283-288.

40. Wang, X., Minasov, G., and Shoichet, B. K. (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* 320, 85-95.

41. Osuna, J., Perez-Blancas, A., and Soberon, X. (2002) Improving a circularly per-muted TEM-1 beta-lactamase by directed evolution. *Protein Eng.* 15, 463-470.

42. Vanhove, M., Raquet, X., and Frere, J. M. (1995) Investigation of the folding pathway of the TEM- 1 beta-lactamase. *Proteins* 22, 110-118.

43. Zahn, R., Axmann, S. E., Rucknagel, K. P., Jaeger, E., Laminet, A. A., and Plückthun, A. (1994) Thermodynamic partitioning model for hydrophobic binding of polypeptides by GroEL. I. GroEL recognizes the signal sequences of beta-lacta-mase precursor. *J. Mol. Biol.* 242, 150-164.

44. Frech, C., Wunderlich, M., Glockshuber, R., and Schmid, F. X. (1996) Competition between DsbA-mediated oxidation and conformational folding of RTEM1 betalactamase. *Biochemistry* 35, 11386-11395.

45. Gill, S. and von Hippel, G. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319-326.

46. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochlo-ride denaturation curves. *Methods Enzymol.* 131, 266-280.

47. Sayle, R. A. and Milner-White, E. J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* 20, 374.

48. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8, 52-56.

49. Santoro, M. M. and Bolen, D, W. (1988) Unfolding free energy changes deter-mined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27, 8063-8068.

索　　引